Title
The dynamics of clathrin-mediated endocytosis in developing embryos

Author
Shirazinejad, Cyna Reza

Publication Date
2023

Peer reviewed|Thesis/dissertation

The dynamics of clathrin-mediated endocytosis in developing embryos

By

Cyna Reza Shirazinejad

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Drubin, Co-Chair
Professor Ian Swinburne, Co-Chair
Professor Dirk Hockemeyer
Professor Srigokul Upadhyayula

Summer 2023

The dynamics of clathrin-mediated endocytosis in developing embryos

By
Cyna Reza Shirazinejad

Abstract

The dynamics of clathrin-mediated endocytosis in developing embryos

by

Cyna Reza Shirazinejad

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor David Drubin, Co-Chair

Professor Ian Swinburne, Co-Chair

Clathrin-mediated endocytosis (CME) is a pathway that remodels the plasma membrane in order to internalize vesicles that contain extracellular cargo such as iron-bound transferrin. Studies of CME have benefited from live-cell imaging where copies of endogenous proteins are visualized by fluorescent protein fusions that are introduced via genome-editing to preserve native protein stoichiometries. Markers of CME such as AP2, DNM2, and ARPC3 can combine in complex and heterogeneous ways, complicating the analysis of data. Here, I describe the development and validation of computational tools that organize and classify different kinds of events marked by the presence of one or more proteins participating in CME. These tools rely on dimensionality reduction to take high-dimensional information (i.e., >100k CME events with highly variable motion profiles and abundance of markers) and compress it into human-interpretable results that describe the kinetics of the markers in their respective groups (Chapter 1). These tools were used to study the dynamics of branched actin network assembly, a force-generating mechanism inside cells, and revealed that sites of stalled CME assemble branched actin networks at one side of the budding CME pit to assist in vesicle internalization (Chapter 2).

Chapter 3 describes the generation of a collection of fluorescently-labeled, genome-edited zebrafish in which diverse complexes and proteins have been tagged: AP2 adaptor complex, Arp2/3 complex, clathrin triskelia, caveolae, and beta-catenin. First, I outline the strategy we used to rapidly generate reagents required for genome-editing. Next, I show that these tagged proteins localize in a living organism in both expected and novel ways by imaging embryos with high-resolution Airyscan microscopy. I first highlight the ability to detect caveolar structures in developing skin

and notochord. Next, I show that we can detect canonical WNT signaling marked by beta-catenin accumulation in the nuclei of several cell types for the first time in a living embryo. The beta-catenin line also revealed potentially novel structures found in neurons and skin. The Arp2/3 complex showed diverse labeling in the form of pegs and ridges during skin wrinkling, including a potentially overlooked apical circumferential ring; punctate labeling in neurons, circulating blood, vascular endothelium, and the swelling endoskeletal disc in the pectoral fin; dense, unresolvable meshes during cloacae morphogenesis; and dense, uniform-appearing accumulation in crawling immune cells. I also observed co-localization of the Arp2/3 complex and AP2 in the pectoral fin, documenting the first evidence for branched actin network assembly use during CME in cells not grown in the artificial environment of a glass surface.

By following the localization of microinjected, labeled transferrin, a key cargo in blood development, I made a surprising observation: transferrin accumulation via CME occurs most abundantly in the vascular endothelium, not in the circulating blood. By providing labeled transferrin at the earliest stages of blood development, I found that emerging red blood cells only consume transferrin much earlier than the onset of circulation. Furthermore, by visualizing the presence of endogenous AP2 and clathrin, I found that emerging blood stem cells have upregulated CME that peaks in frequency above the nearly non-existent levels found in the cells that precede them, the dorsal aorta endothelium.

My zebrafish data demonstrated the utility of studying endogenous markers of cellular pathways that highlight the known use of these proteins as well as producing new insights into their use.

Dedicated To
Anna, Mom, Dad, Cassra, Ali, Sokka, and Zuko

Table of Contents

List of Figures and Tables

Common abbreviations:

- AO-LLSM: adaptive optical lattice light-sheet microscopy
- AP2: Adaptor Protein 2
- ARPC3: actin related protein 2/3 complex subunit 3
- BIC: Bayesian Information Criterion
- CCP: clathrin-coated pit
- CCV: clathrin-coated vesicle
- CME: clathrin-mediated endocytosis
- CWS: canonical WNT signaling
- DNM2: Dynamin-2
- ECDF: empirical cumulative distribution function
- F0: Crispant embryo, potentially Founder generation
- F1: generation recovered from F0 out-cross
- FOV: field-of-view
- GMM: Gaussian Mixture Model
- PCA: principal component analysis
- PSF: point spread function
- SNP: single nucleotide polymorphism
- TIRF: total internal reflection microscopy
- ap2s1: Adaptor Protein 2 sigma subunit
- cav1: caveolin 1
- clta: clathrin light chain A
- ctnnb1: beta-catenin 1
- het: heterozygous
- homo: homozygous
- hpf: hours post fertilization
- ix: in-cross
- mNG: moxNeonGreen
- mSI: mScarlet-I
- ox: out-cross

Acknowledgements

Thanks party people

Chapter 1: data mining tools for separating tracked signals from noise

Background and Significance

In this chapter I describe tools I developed to aid interpretation and analysis of the dynamics of a process called clathrin-mediated endocytosis. Clathrin-mediated endocytosis (CME) is a cellular pathway that remodels the plasma membrane into endocytic vesicles[1]. CME can be harnessed to move receptors and their extracellular cargos into the cell through formation of endocytic vesicles. Thus, CME is a selective portal of entry from the outside of the cell into the cell. Much of our understanding of the dynamics of proteins in clathrin-mediated endocytosis has been the result of real-time fluorescence imaging[2].

Fluorescent movies of live cells can contain a wealth of information[3]. From a data science point-of-view, they present unique challenges: diverse patterns of labels, complex dynamics, multimodal distributions, and a large dynamic range of phenotypes associated with perturbations. A central goal of cell biology is visualizing the position and movement of organelles in order to decipher the rules for how they are built, maintained, and disassembled[4]. Yet, there has been a lag in interpretable and unbiased computational tools that allow us to extract meaningful information from image-based technologies advancing at a rapid pace that can provide otherwise unattainable mechanistic insights[5]. Broadly speaking, the tools I developed, which are described in this chapter, aim to address the following questions: How heterogenous are my data? How are the proteins-of-interest moving together? What are the categories of coordinated protein movement within my data?

Fluorescence microscopy has resulted in an explosion in the amount of data being generated in studies of CME[2]. Fluorescence movies can be attained by expressing proteins of interest as fusions to protein fluorophores like Green Fluorescent Protein (GFP). To achieve this, the coding sequence of a protein of interest is fused with the DNA encoding a fluorescent protein. These chimeric DNA constructs can be expressed by cells so the protein-of-interest is labeled with a fluorescent protein. If done by genome editing the chromosomal copy of a gene, physiological protein expression levels can be maintained. This approach to protein labeling can be multiplexed, wherein multiple proteins of interest are labeled with orthogonally-colored fluorophores, allowing the dynamics of multiple proteins to be visualized. Advances in live-cell imaging have allowed fast, high-resolution, and gentle imaging of cells, providing glimpses into how molecules move within cells[6].

Another advantage of this approach is that it can be made quantitative to reveal CME site-to-site heterogeneity[7]. Analysis of site-to-site variation provides a powerful avenue toward mechanistic insights by revealing co-variance and anti-correlation of paired markers and parameters. To make these fluorescence studies quantitative, it is imperative that proteins be studied at their endogenous stoichiometry. The dynamic range of copy numbers of CME proteins ranges from a few proteins (~5) to hundreds of each protein per site, with well over 50 different proteins having documented

participation in CME[8,9]. Therefore, to image a cells' steady-state protein expression, the labeling of endogenous proteins must be achieved.

This approach of studying genome-edited fusion proteins expressed at endogenous levels in live cells has been a staple in studies of yeast endocytosis, where yeast molecular-genetics allows for simple and rapid integration of exogenous DNA seamlessly into the endogenous genomic locus for a protein of interest[10]. Genome editing results in, for example, fluorescent protein-encoding DNA fused to the coding sequence for the carboxyl-terminus of a target protein. Once a time-limiting step in studies of metazoan cells, precise integration of exogenous DNA (e.g. fluorescent protein) at endogenous genomic loci has been made fast and efficient by advances in genome-editing technology. Genome-editing has benefited studies of cell biology by providing a cheap and fast method for precise editing of genomes[11]. Advances in editing technologies have allowed cell biologists to survey many cellular pathways, resulting in mechanistic studies of how the visualized proteins promote the pathway's success[12]. Paired with a rapidly evolving field of live-cell microscopy, rich datasets can rapidly be generated that allow for reformed views of how proteins in cells are organized dynamically. Nevertheless, fluorescent movies often contain data that is difficult to parse and synthesize into interpretable phenotypes.

The CME field in particular has benefited from particle-tracking tools[13,14]. In the context of watching CME, particle-tracking provides information about the genesis and disappearance of CME sites. Proteins recruited to CME sites assemble over a span of roughly one-two minutes. For example, when viewing cells via Total Internal Fluorescence Microscopy (TIRF), which excites fluorophores near the "bottom" of cells, spots appear and disappear[15]. TIRF is an excellent choice for studying CME because it is highly selective for fluorophores at the bottom plasma membrane of cells (ignoring fluorophore activity "inside" the cell) and because it is very sensitive, making possible high-time resolution image acquisition[16].

When tracked through time, CME spots grow in size as a population of labeled protein assembles at a particular CME site, and eventually disappears. This disappearance can be roughly explained in two ways: (1) the disassembly of the protein from the site, or (2) from "upward" movement of the CME protein complex out of the evanescent wave that activates the fluorophores (i.e., the CME site is physically displaced from the plasma membrane during membrane invagination and vesicle release). Combined with the knowledge that proteins at CME sites are organized into roughly-circular assemblies 100 nanometers in diameter (smaller than conventional TIRF resolution), CME sites look like spots that start off as faint flickers, and then grow brighter and brighter before eventually disappearing from the movie.

A useful particle-tracking system will detect the first appearance of a CME site, keep track of multiple colors (i.e., protein imaging channels), follow the site as it wiggles around (e.g., cell movement or intra-membrane diffusion), and mark its time and position of disappearance. Thus, a "track" is, constrained by optical resolution, a dynamic spot that is a temporal vector containing fitted positions and fluorescence

amplitudes. In the case of TIRF, a 2-D projection of fluorescence activity near the bottom surface of cells, track coordinates are (X, Y) coordinates and brightness amplitudes that are a function of the time-points during which the spot is surveyed.

Much of our lab, as well as the field studying CME, has relied on a program called cmeAnalysis: a MATLAB program from the Schmid and Danuser Labs (University of Texas Southwestern Medical School)[13]. cmeAnalysis contains several modules: a particle detection algorithm that finds spots, a tracking algorithm that links spots found in temporally adjacent images, and a selection algorithm that classifies tracks based on the significant presence of two or more markers. The cmeAnalysis suite offers excellent performance on detection and tracking. However, as my work illustrates, upon closer inspection, the classification scheme struggles to unmix a heterogeneous population of tracks. A central goal of my work is to untangle various populations of moving spots and to make possible extraction of as much information from CME data as possible.

I was lucky to be in the Drubin lab at the same time as Dr. Meiyan Jin. Motivated by a work-from-home situation and a wealth of data she had generated, the formative and early phase of my PhD was in collaboration with her to build better computational tools for studying CME. Building these computational tools opened doors to studies beyond our initial line of questioning. We started with a handful of questions that kickstarted the tool development used in Chapters 2 and 3. These tools answered those questions then did what every computational tool should do: they stimulated new lines of questioning and removed human bias to the fullest extent possible. New experiments were inspired and the subsequent analyses were rendered relatively painless.

Meiyan had generated three stem cell lines that paved the foundation of our work. A few years prior, stem cells became the main mammalian cell biology workhorse in the Drubin Lab. Human induced pluripotent stem cells (hiPSC) had a few notable benefits over cancer-derived cell lines: a normal number of chromosomes (i.e., genomic stability), they were derived from a healthy human donor (not from tumors like most tissue culture lines), and they were readily differentiated in culture into cell types from various tissues[17]. The lab had an unpublished hiPSC line generated by a previous member of the Drubin Lab, Sun Hae Hong, in which a subunit of the AP2 complex was tagged with a red fluorescent protein, TagRFP-T. The AP2 complex is among the early arriving proteins in CME and is a suitable marker for CME because it gets wedged between the plasma membrane and the clathrin coat, making it almost omnipresent for the duration of CME[18].

Meiyan derived the other three lines from the AP2-TagRFP-T line. All three lines contain TagGFP2 fusions with dynamin-2 (DNM2), an important marker for vesicle scission[7]. A burst of DNM2 assembly is correlated with successful endocytosis. DNM2 appears early in endocytosis in low levels, but its further assembly is rapid when DNM2 rings assemble at the neck of omega-shaped invaginations as shown in Figure 1.1[19]. The AP2-TagRFP-T/DNM2-TagGFP2 (aka AP2/DNM2) line is the “base” reference for CME dynamics that is used to identify clathrin-coated pits (CCPs), which are defined later. To study branched actin network assembly and its influence on CCP maturation, Meiyan

engineered two other hiPSC lines: AP2/DNM2/ARPC3-HaloTag and AP2/DNM2/N-WASP-HaloTag. ARPC3 is a subunit of the Arp2/3 complex[20]. A branched actin network is, at an ultrastructural level, a dynamic network of dendritic appearance that assembles actin filaments from actin monomers that can push and pull on membranes. The branching of actin filaments--linear rods of actin monomers--is mediated by the Arp2/3 complex which binds to a “mother” filament and nucleates a “daughter” filament. Branched actin networks are known to mostly have a facilitatory role in mammalian cells’ endocytosis, which is in direct contrast with its strict requirement for yeast endocytosis[21,22]. My work with Meiyan set out to better understand the dynamics of the Arp2/3 complex at CME sites in mammalian cells. The AP2/DNM2 cell line and AP2/DNM2/ARPC3 cell lines will be denoted as the 2 and 3 color cell lines, respectively. AP2, DNM2, and ARPC3 together mark three endocytic modules starting from coat initiation, vesicle scission, and cytoskeletal force-generation at the plasma membrane[15].

Meiyan came to me with two general questions about her data: First, could we automatically detect spots with “strong enrichment” for AP2 and DNM2 that cmeAnalysis incorrectly grouped with those that had “weak enrichment”? Second, were spots detected in one channel, in this case ARPC3, actually *slightly* separated spatially from AP2 and DNM2? Addressing the first question created a foundation for categorizing tracks in a way that allowed us to parse through mounds of data that followed after my initial model prototyping and validation, outlined in Chapter 2. The second question was a valuable lesson for me in creating critical tests that highlight the limits of light-microscopy that can cause mischaracterizations of colocalization resulting from chromatic aberration and diffusing puncta. These tests are important because they address and dismiss the alternative explanations for how microscopy artifacts could create the effects we observe.

Figure 1.1: Overview of clathrin-mediated endocytosis

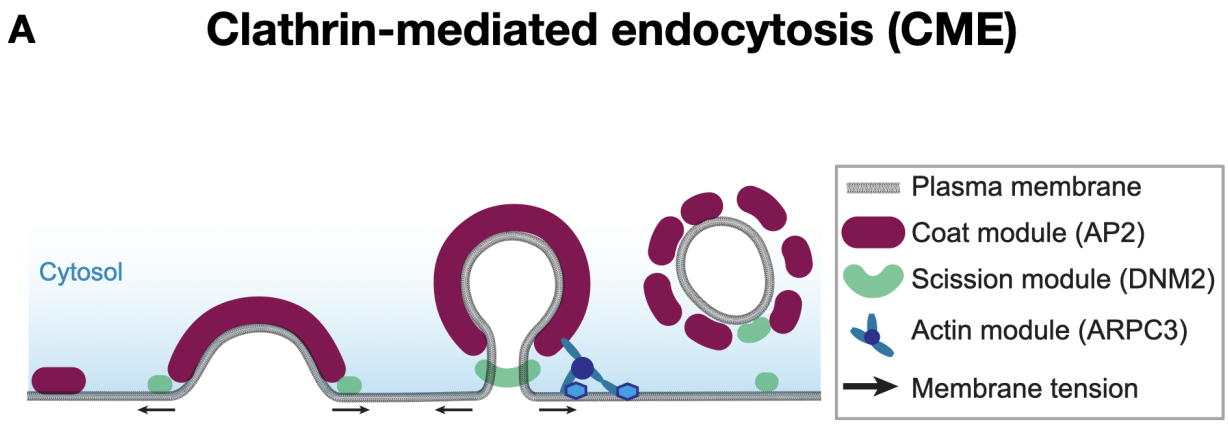


Figure 1.1 legend: A) This cartoon represents a 2D slice the plasma membrane of a cell, with the “top” of the cell pointing up and “bottom” of the cell pointing down, with the gray stripe representing the basal plasma membrane. Clathrin is not depicted because it

is not tagged. Starting from the left, a flat piece of plasma membrane is remodeled into a vesicle. As time progresses from left to right, the coat module associates with the inner leaflet of the plasma membrane. Eventually, sufficient curvature is generated in the neck of the “omega-shaped” invagination to recruit the scission module in a ring that wraps the neck. Force generation that counteracts membrane tension through the actin module frees the vesicle from the plasma membrane.

Results

AP2 was used as the fiducial marker for clathrin-mediated endocytosis (CME) in all of our tracking experiments in cmeAnalysis. DNM2 was used as a secondary channel to mark vesicle scission and the termination of vesicle formation. cmeAnalysis provides a classification for the secondary tracking channel, DNM2, as significant or insignificant. This designation determines whether the detection of DNM2 would have been registered as an independent event without its detection being dependent on the primary marker, AP2, being detected and tracked through time. The model used for determining DNM2’s significance does not require that DNM2 detections be consecutive or be present in specific stages of AP2’s lifetime. However, previous studies have indicated that DNM2 is recruited at low levels in the early development of the endocytic site as well as in a rapid burst of recruitment just prior to scission. To generate the models in this chapter, the 2-color (AP2/DNM2) line was used. All analyses were performed in a set of 20 Jupyter Notebooks accompanied by a Python library for routine calculations.

Several previous attempts have been made to distinguish between vesicles visiting the TIRF field and authentic clathrin-coated pits (CCPs) and vesicles originating at the plasma membrane[9,13,18]. These studies used a combination of hard-percentile thresholds of measured intensities, specific hand-engineered features, or supervised machine learning techniques. Our methods were motivated by a desire to increase the range of features that can be used to select authentic CCPs without the use of single features or arbitrary thresholds in lifetimes and intensities. Phenotypic measurements of CME events typically include the lifetime and relative brightness of measured intensities of tracked events over time.

Feature decomposition and clustering

The advantage of automated particle tracking is that it allows further characterization of tracked events. Our aim was to define observables that could be extracted from fitted positions and amplitudes of tracked events that describe events in an abstraction that simplifies amplitudes and positions as a function of time, A(t) and (x(t), y(t)), respectively. The observables, or features, are separated into seven modules. The ‘brightness’ module describes each tracks’ lifetime (marked by AP2 detections in the primary channel) and maximum intensities of AP2 and DNM2. Subsequent modules describe the motion of events, position of peak intensities, rates of intensity changes through time, relative peak characteristics between channels, signal moments or shapes, and frame-by-frame detection significance of the secondary channel (DNM2).

These parameters allow a more careful understanding on how potential changes to CME dynamics can arise from perturbations such as drug treatments, plasmid expression, and knock-downs. Table 1.1 summarizes the 32 engineered features.

Table 1.1: features used for track decomposition, per track

note: intensities and positions are fitted, not raw values taken from fluorescence micrographs

feature number	feature name	feature description
1	lifetime	time between the first and last frame of detected AP2 (seconds)
2	max_int_ap2	maximum intensity AP2 (a.u. fluorescence)
3	max_int_dnm2	maximum intensity DNM2 (a.u. fluorescence)
4	dist_traveled_ap2	track start-to-finish net movement (pixels)
5	dist_traveled_dnm2	track start-to-finish net movement (pixels)
6	max_dist_between_ap2_dnm2	the maximum frame-to-frame separation between AP2 and DNM2 (pixels)
7	md_ap2	mean displacement (pixels)
8	md_dnm2	mean displacement (pixels)
9	time_to_peak_ap2	time for the intensity to reach its peak (seconds) [0 if peak is first frame]
10	time_to_peak_dnm2	time for the intensity to reach its peak (seconds) [0 if peak is first frame]
11	time_after_peak_ap2	time for intensity to decay from its peak (seconds) [0 if peak is last frame]
12	time_after_peak_dnm2	time for intensity to decay from its peak (seconds) [0 if peak is last frame]
13	time_between_peaks_ap2_dnm2	time between peaks of two channels (seconds)
14	avg_int_change_to_peak_ap	average change in intensity to the peak (a.u.

	2	fluorescence) [0 if peak is first frame]
15	avg_int_change_to_peak_dnm2	average change in intensity to the peak (a.u. fluorescence) [0 if peak is first frame]
16	avg_int_change_after_peak_ap2	average change in intensity after the peak (a.u. fluorescence) [0 if peak is last frame]
17	avg_int_change_after_peak_dnm2	average change in intensity after the peak (a.u. fluorescence) [0 if peak is last frame]
18	peak_int_diff_ap2_dnm2	difference between maximum intensity of channel 0 and channel 1 (a.u. fluorescence)
19	ratio_max_int_ap2_dnm2	ratio between maximum intensity of channel 0 and channel 1 (unitless)
20	mean_ap2	average of fluorescence (a.u. fluorescence)
21	mean_dnm2	average of fluorescence (a.u. fluorescence)
22	variation_ap2	variation of fluorescence (a.u. fluorescence^2)
23	variation_dnm2	variation of fluorescence (a.u. fluorescence^2)
24	skewness_ap2	skewness of fluorescence (unitless)
25	skewness_dnm2	skewness of fluorescence (unitless)
26	kurtosis_ap2	kurtosis of fluorescence (unitless)
27	kurtosis_dnm2	kurtosis of fluorescence (unitless)
28	number_significant_dnm2	number of significant detections with p-val (Anderson-Darling normality test) lower than provided threshold (counts) [p-value < 0.01]
29	max_consecutive_significant_dnm2	maximum number of consecutive significant detections with p-value lower than provided threshold (counts) [p-value < 0.01]
30	fraction_significant_dnm2	fraction of event with significant detections with p-value lower than provided threshold (unitless) [p-value < 0.01]
31	fraction_peak_ap2	fraction of the event where the peak is located (unitless)

32	fraction_peak_dnm2	fraction of the event where the peak is located (unitless)

All AP2/DNM2 cell line tracks are first uploaded into Python Jupyter Notebooks as “ProcessedTracks.mat”. cmeAnalysis categorizes tracks into 8 groups. The first four consist of single events and the last four consist of splitting/merging events. Category 1, or ‘valid”, tracks are those that have gaps that do not exceed the designated ceiling of 2 consecutive gaps. Valid tracks also show up after the movie begins and leave before the movie ends. Category 2 tracks have excess gaps. Category 3 tracks are either present from the start of the movie or are present at the end of the movie. Category 4 tracks are present throughout the entire movie. Categories 5-8 are similar to category 1-4 tracks except they split and/or merge. Only valid tracks are considered for the remainder of the analysis to work with the cleanest and most interpretable data. Figure 1.2 A shows that the size of puncta (measured by the standard deviation of the positional fit) is tightly distributed across different experimental fields-of-view (FOV). We did not perform statistical analysis of these distributions because we did not discriminate between datasets using this information. However, it is clear that the “orange” dataset has slightly larger puncta which could be the result of acute differences in focus, particle diffusion, or larger clathrin-coated structures found in the plasma membrane of those cells. ~99% of all tracks from every FOV have a width less than ~2 pixels (2 standard deviations on either side of the mean) or ~200 nanometers in size. The larger size of these detection structures can likely be attributed to the long exposure time (300 ms) that integrates the intensity of smaller mobile structures that blur during camera integration. Figure 1.3 A shows that ~50% of tracks are valid. These tracks are kept for further analysis. Roughly 6,000 to 8,000 valid tracks (Figure 1.3 B) are kept from each FOV.

Figure 1.2: fitted width of puncta found in valid tracks, per FOV

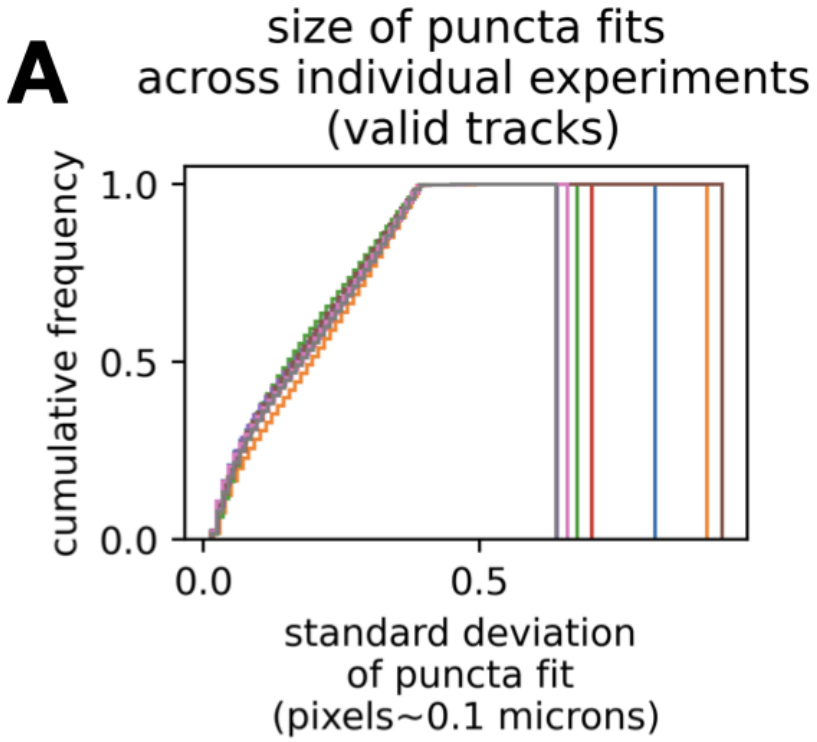


Figure 1.2 legend: A) the empirical cumulative distribution function (ECDF) of the pooled fits to puncta size per track. Each line represents a FOV, and every track has its time-points pooled into the distribution for the FOV.

Figure 1.3: categories of tracks detected by cmeAnalysis, per FOV

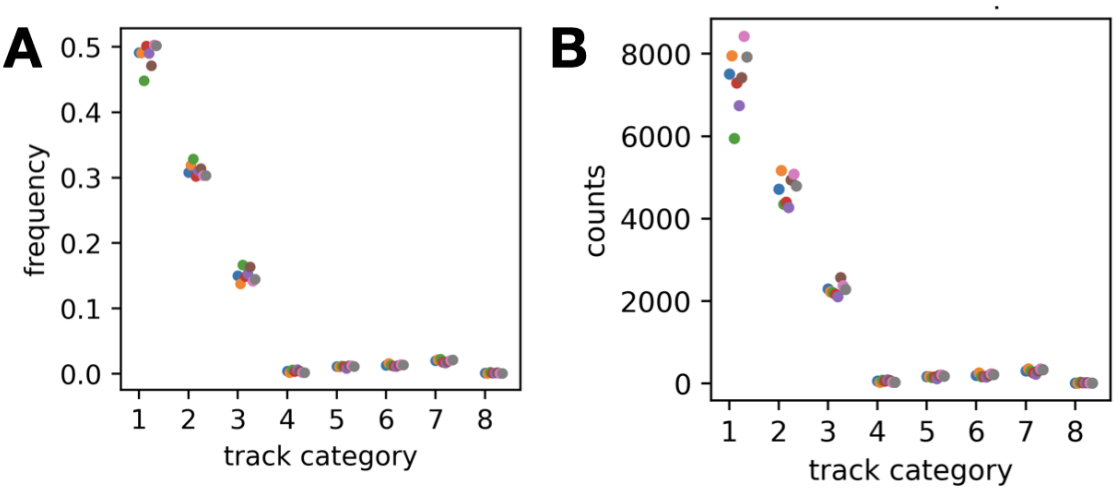


Figure 1.3 figure: The breakdown of cmeAnalysis categories per FOV for all detected tracks. In A, categorization is by frequency and in B, categorization is by raw counts.

Each valid track is decomposed into the 32 features previously described using an object-oriented approach to select the desired features needed for each track. Then, in

addition to these features, each track is additionally labeled with the number of channels in the cell line (2), the imaging date, and cmeAnalysis' prediction of DNM2-positive or negative. From here on, cmeAnalysisDNM2+ or cmeAnalysisDNM2- will be used to reference tracks identified by cmeAnalysis that did recruit DNM2 and those that did not recruit DNM2, respectively.

We next wanted to discover natural patterns of similarly behaved events within the feature space. The goal of this is to extract events that fall within distinct phenotypic categories that may reveal functional differences. Most importantly, we are now seeking tracks that have a characteristic DNM2 peak that corresponds to vesicle scission. To do so, we target events that have the most DNM2 recruitment relative to the entire population of tracks.

While the raw feature distributions showed that tracks could possibly fall within two or more modes, there were no clear tails between populations of events (Figure 1.4). Furthermore, with 32 dimensions in the feature space, pairwise feature comparisons and boundary drawing by hand to generate thresholds becomes tedious and will introduce arbitrary decision making.

Figure 1.4: raw feature histograms

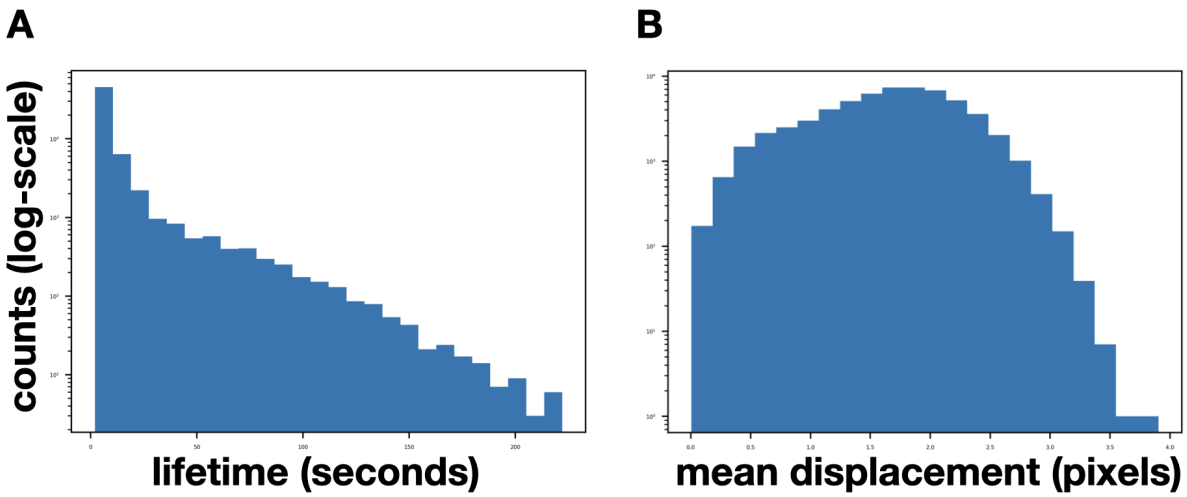


Figure 1.4 legend: Raw features of two commonly used features for threshold-based event discrimination. A: coat lifetimes (AP2) and B: mean frame-to-frame puncta displacement.

To make the visualization of similarly behaved events possible, we turned to linear dimensionality reduction with Principal Component Analysis (PCA). The relative scale of the features between one another are disparate. Therefore, it is imperative to scale features to alternative distributions for ease of comparison. We attempted a suite of scaling options and found, when viewing the first two principal components, that scaling each feature to normal and uniform distributions revealed distinct modes. This work highlights the importance of choosing an appropriate feature scaling method for

exploratory data analysis. I suspect the difficulty of this problem is due to heterogeneity in feature distributions that causes insufficient separation without normally-distributed quantile rescaling.

The first two principal components of the normal-scaled data revealed a minimum of 5 clear modes when viewed in log scale: clear peaks in event densities were surrounded by events with decreasing densities. Additionally, the cumulative explained variance (Figure 1.5) for normally-scaled features shows a smooth increase with additional principal components (PC). For interpretation's sake, we only analyzed the first two PCs, which captures ~50% of the variance with two components. Additionally, there was not a clear "elbow" when looking at the right axes (Figure 1.5 B) to see if additional components retrieved large changes in the fraction of cumulative explained variance of all features. Further work is required to analyze higher-dimensional modes found in additional PC axes.

Figure 1.5: cumulative explained variance versus number of PCA components, per scaling option

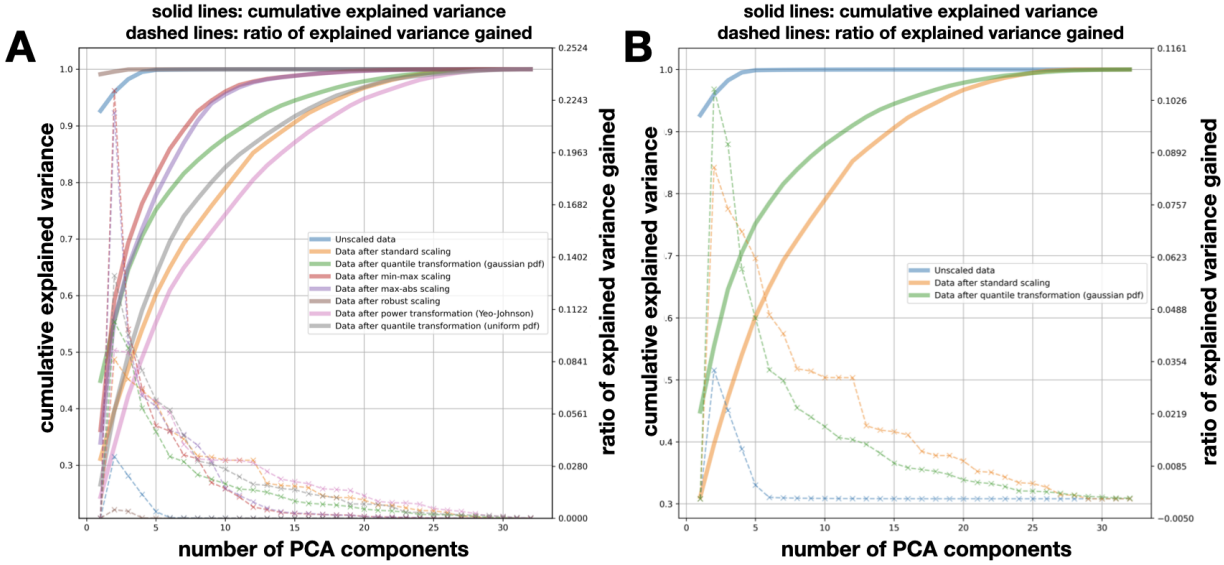
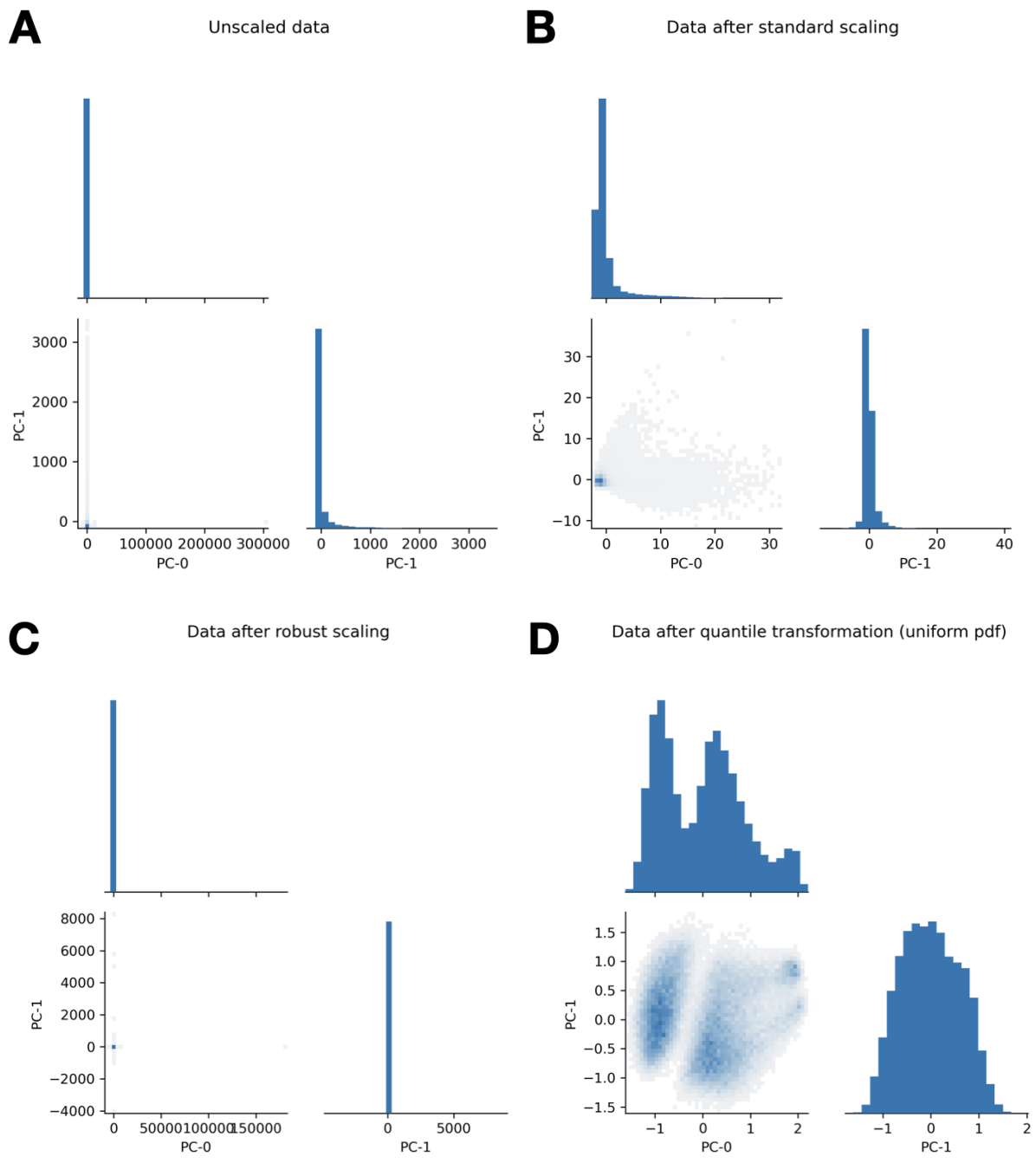


Figure 1.5 legend: A: left axis/solid lines: cumulative explained variance as a function of number of PC components per feature scaling option. Right axis/dashed lines: change in cumulative explained variance with an additional component. B: Repeat of A with only unscaled data, standard scaling, and normally-distributed quantile transformation (used for subsequent modeling).

Figure 1.6: first two principal component 2-D histograms, per scaling option



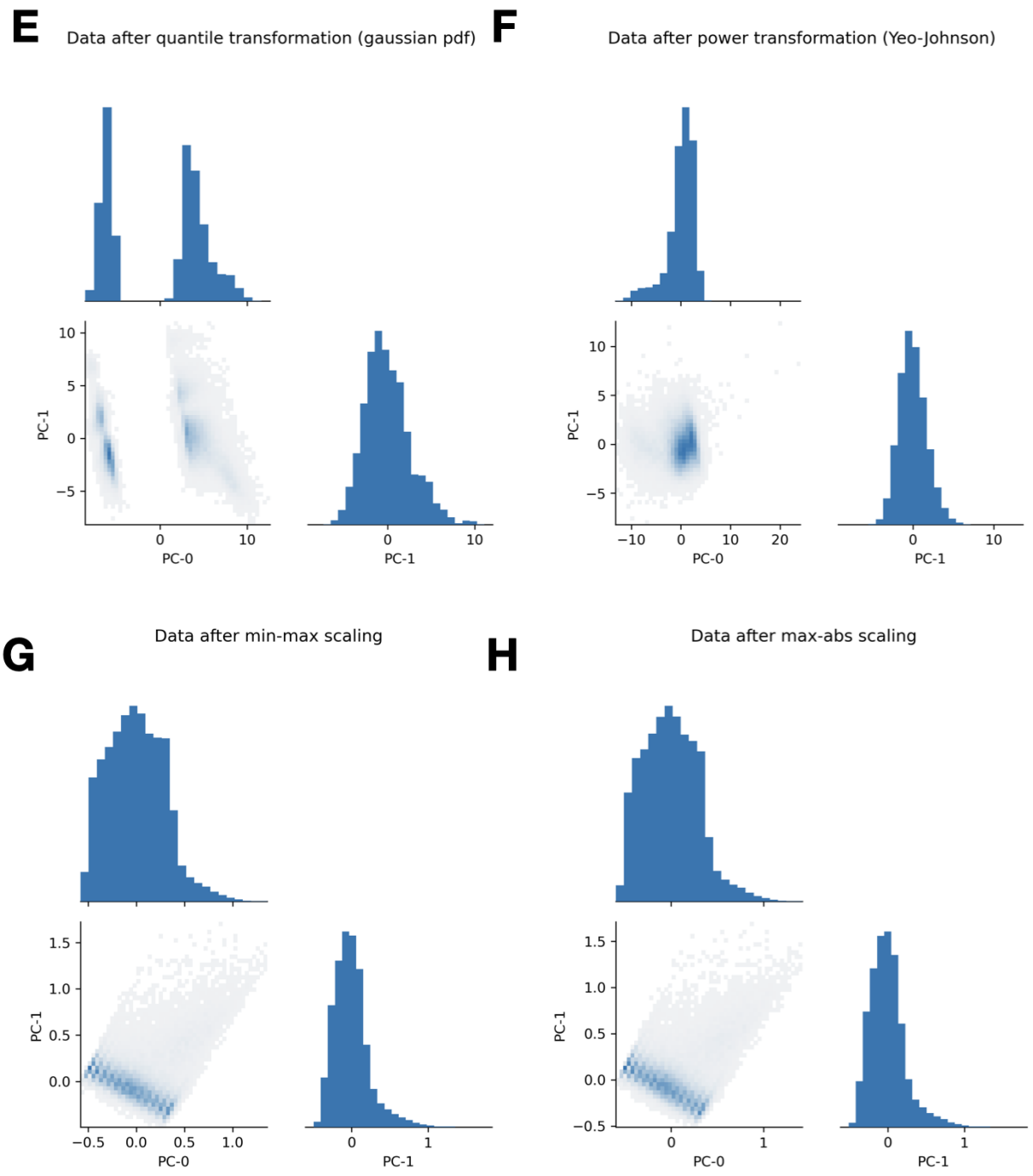


Figure 1.6 legend: 2-D density histograms of the first two principal components.

We used a Gaussian Mixture Model (GMM) to cluster events into different groups. While GMMs are an algorithm traditionally used for density estimation, they provide an advantage here over k-means clustering since there are soft boundaries and allow for non-diagonal covariance matrices to be computed. There was an elbow at 5 components when measuring the Bayesian Information Criterion (BIC) as a function of additional GMM components. The BIC describes the trade-off in the goodness of fit

gained by additional components. A lower BIC represents a better fit model. As more components were added in the GMM, the BIC continues to decay and does not eventually bend upwards, which might indicate overfitting. This conclusion was true in the aggregate dataset model (all 8 FOVs merged) as well as individual FOVs (inset in Figure 1.7). However, since the BIC starts to decay less quickly at 5 components, a 5 component model was chosen. Cluster identities were assigned based on hard clustering corresponding to the highest probability of each event to the candidate clusters (Figure 1.8). While it is visually clear that there are additional modes in the first two PC density plots (Figure 1.8), we proceeded with cluster “0”, which corresponds to the events with the strongest DNM2 enrichment.

Figure 1.7: Bayesian Information Criterion versus number of Gaussian Mixture Model components, aggregate and per FOV

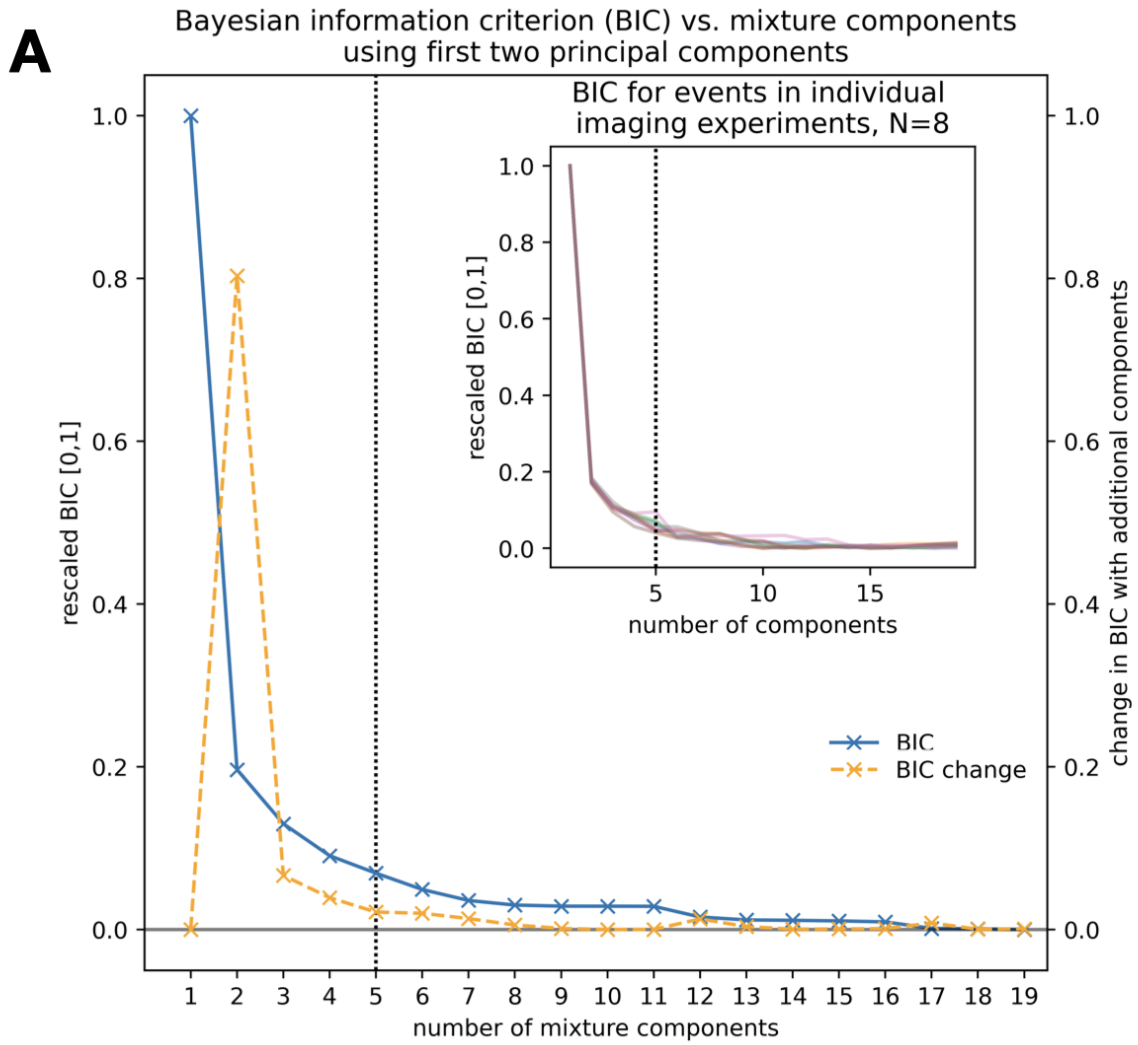


Figure 1.7 legend: left axis: The BIC for the aggregate model was rescaled from 0-1 and showed a decay profile with additional GMM components. The change in rescaled BIC component-to-component showed a gradual decline that approached ~0 around ~5-9 components in.

Figure 1.8: 2-D histogram of normally-distributed feature principal component projection, aggregate datasets

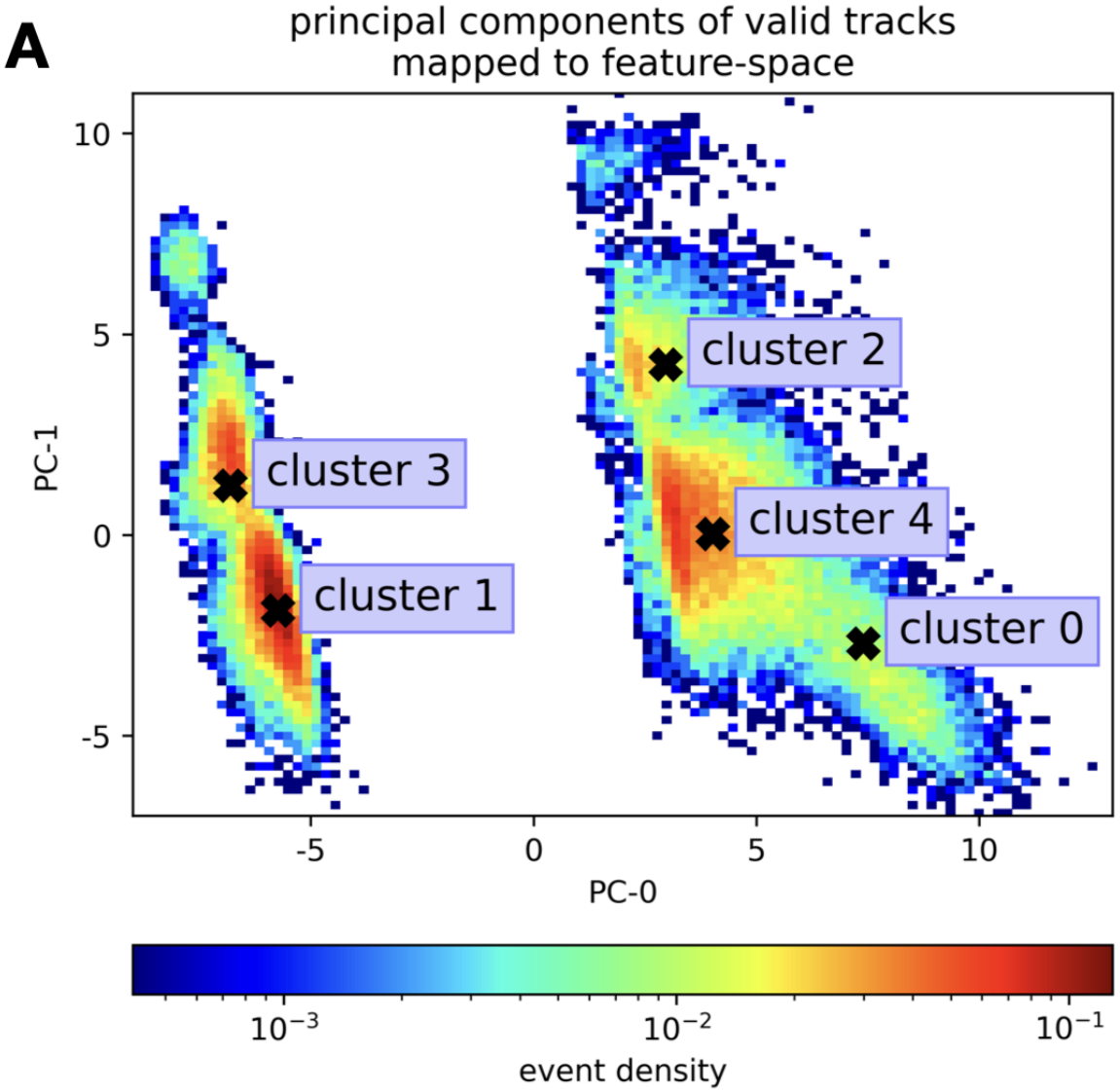


Figure 1.8 legend: 2-D density histogram of first events in principal component space with annotated cluster centers from GMM

By simulating data points on a grid that ranges x and y values near the bounds of the empirical PC-0 and PC-1 scatter points, we were able to see the sharpness of the clustering by subjecting these simulated points to posterior inference from the GMM. Most of the events inside clusters had a posterior probability of >90% association with their nearest cluster (yellow background). However, events situated in between clusters had posterior probabilities of 40-80% that required further interrogation (Figure 1.9).

Figure 1.9: PCA clusters overlaid with simulated data points in background

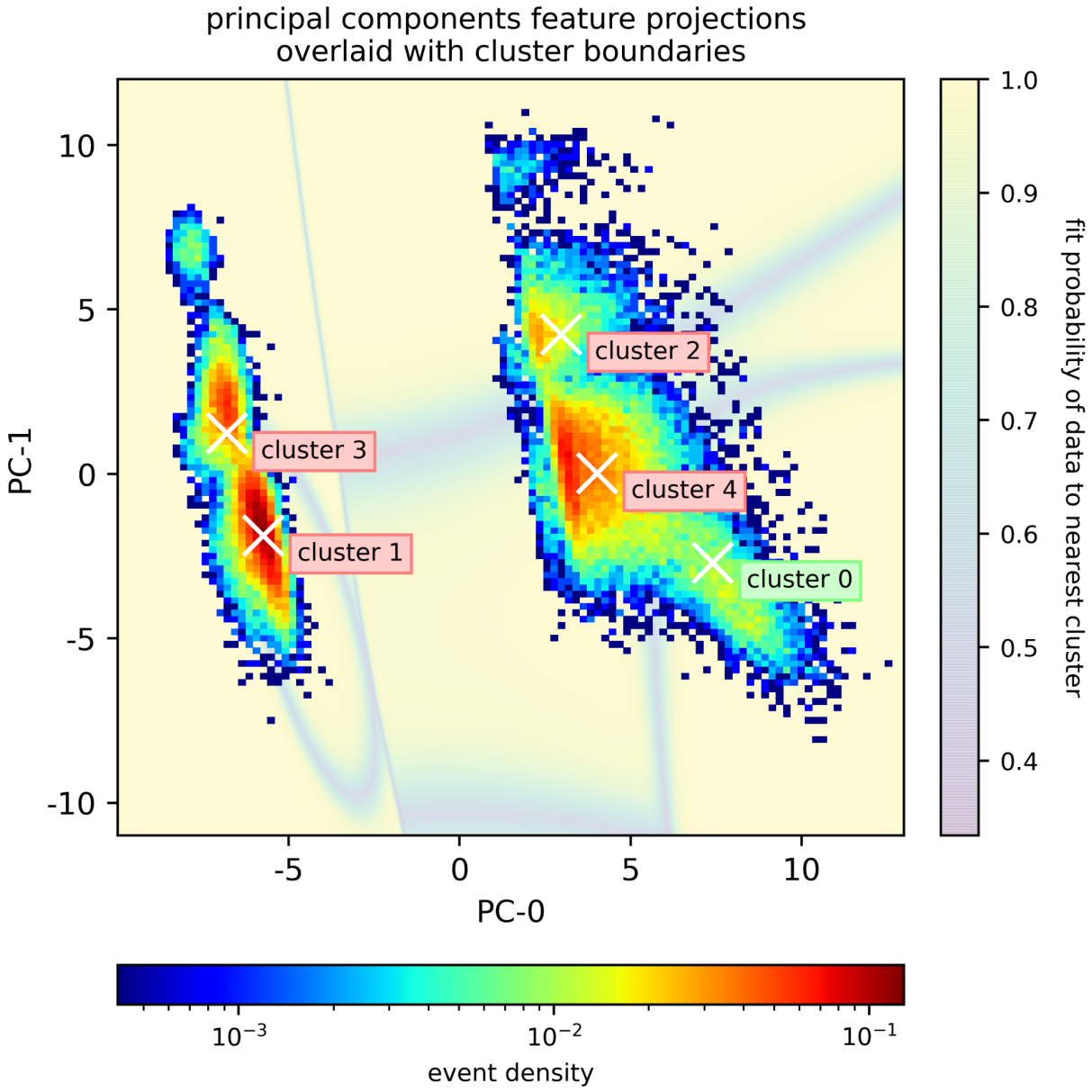


Figure 1.9 legend: foreground: 2-D density histogram of first events in principal component space with annotated cluster centers from GMM, background: simulated data points on grid with colors corresponding to posterior probability of hard clustering assignment from 5 component GMM.

The principal components of the normal-scaled data could be related to the feature space from which the maximal variance directions are computed (Figure 1.10). We found that the directions of maximal variance in the data corresponded to lifetimes, maximum intensities, motilities, and relative DNM2 recruitment concomitant with AP2. Features such as lifetimes, intensities, and motilities were used to generate discriminating boundaries between tracked events to discriminate 'authentic' CME events from putative 'visitors'[14,18]. These results demonstrate the ability to separate

events without assumptions about the degree of DNM2 association with AP2. While these data show that most of the 32 features are not important for generating PC modes, we did not have to assign any thresholds between the dominant features to cluster events.

Figure 1.10: magnitude and directions of first two principal components

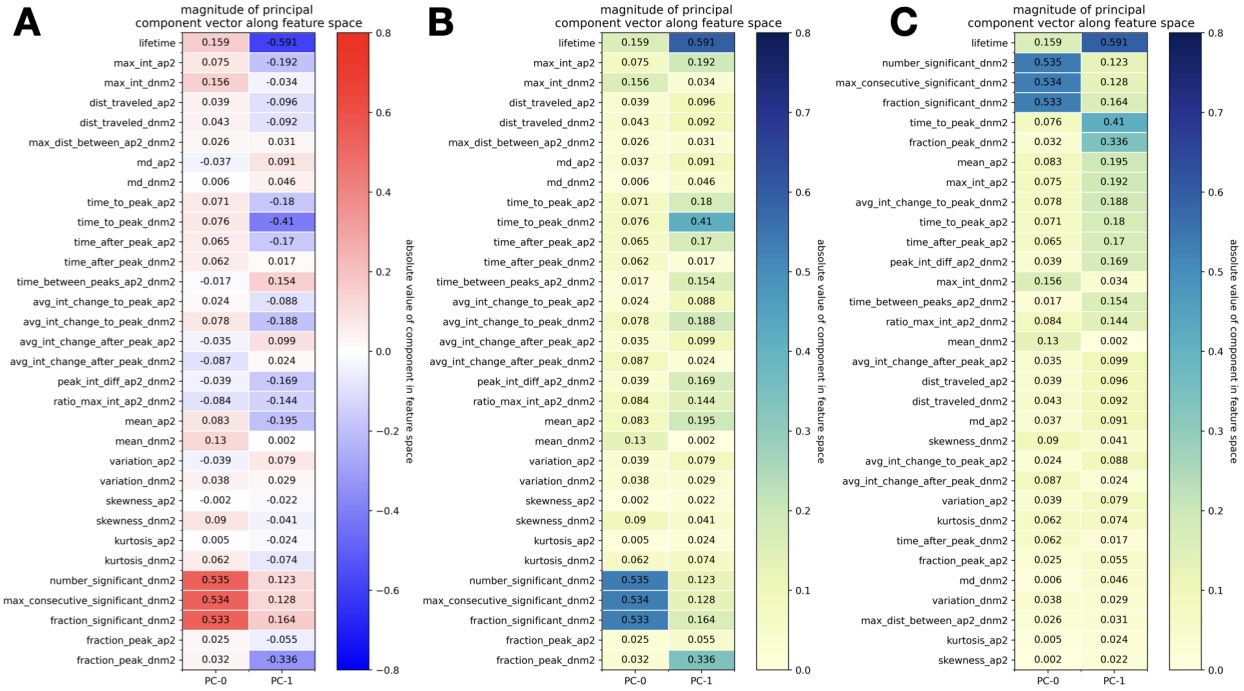


Figure 1.10 legend: A: The weights and directions of principal components along feature space. B: Magnitudes of principal components along feature space. C: B but sorted vertically.

Additionally, when viewing the cmeAnalysis binary classification of cmeAnaylsisDNM2+/- events, we found 34% of cmeAnalysis+ events to be spread over all 5 clusters (Figure 1.11). The majority of cmeAnalysis+ events shown in Table 1.2 correspond to cluster 4. Our model is able to break down the cmeAnalysis' binary classification of secondary co-localization channels (DNM2) into further groups.

Figure 1.11: relationship between cmeAnalysis+ events and PC modes color coded by cmeAnalysis+/-

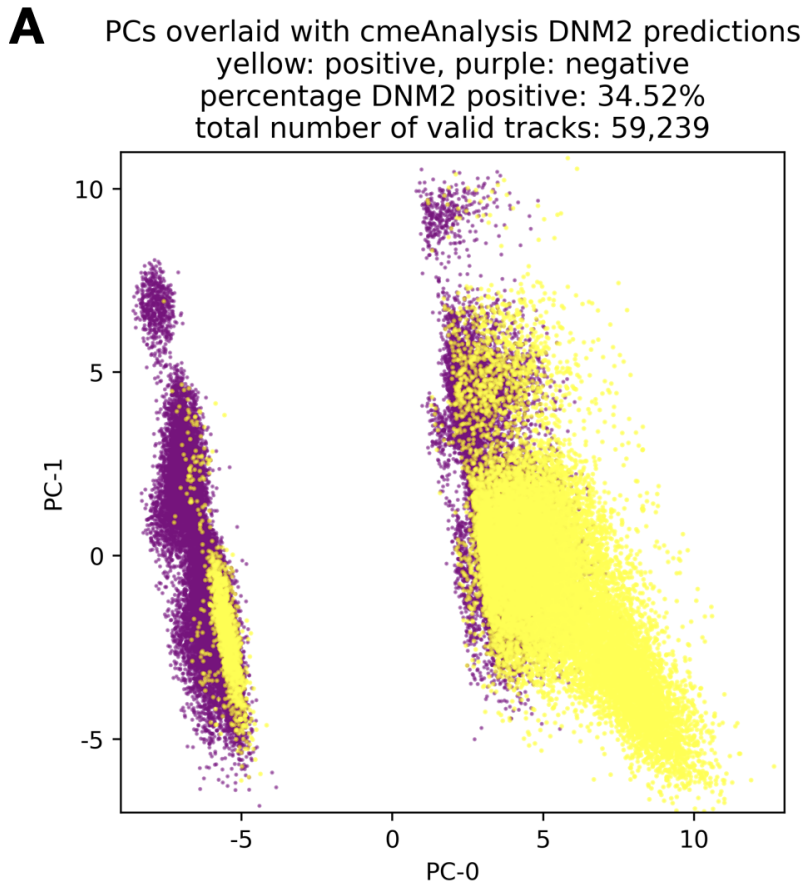


Figure 1.11 legend: scatter plot of valid tracks in PC space with colors denoting cmeAnalysis+ (yellow) and cmeAnalysis- (purple) binary classification

Table 1.2: overlap of PCA-based clusters with binary DNM2 classification from cmeAnalysis

PCA cluster	number of tracks	number of tracks predicted to be DNM2- by cmeAnalysis	number of tracks predicted to be DNM2+ by cmeAnalysis
0	5938	50	5888
1	14571	13134	1437
2	6181	5014	1167
3	10091	9982	109
4	22458	10612	11846

The highest probability fit of each event to each cluster was used to determine the initial assignment of each event into GMM clusters. Then, the cluster with the highest average of maximum DNM2 intensities was selected to identify the preliminary DNM2 positive candidates in cluster 0 (DNM2+ from hereon).

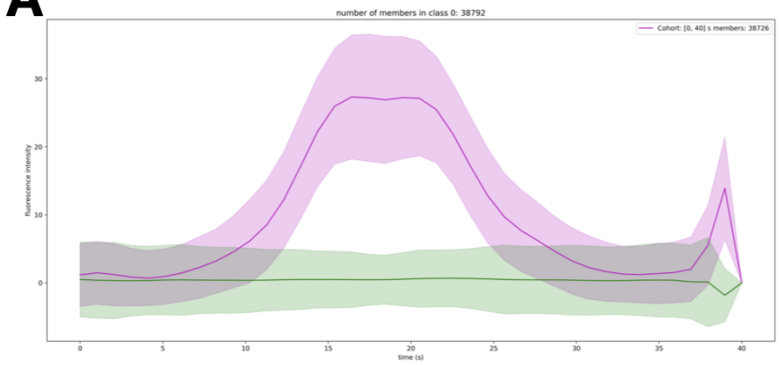
To aid in the interpretability of the events in different GMM components, tracks were aligned within cohorts for visualization of their average behavior. Here, the alignment procedure follows that of cmeAnalysis' algorithm closely. First, tracks are binned within lifetime cohorts: <40, 40-60, 60-80, and >80 seconds with inclusive lower bounds and exclusive upper bounds. No modifications are made from cmeAnalysis besides interpolating to the ceiling of the cohort bounds rather than the mean of the upper and lower bounds. Briefly, intensity traces for each track are interpolated to the ceiling of the cohort and averaged across all time domain points in the interpolated space. For all instances of cohort plotting, 0.25 standard deviations above and below the mean are displayed with transparent hue.

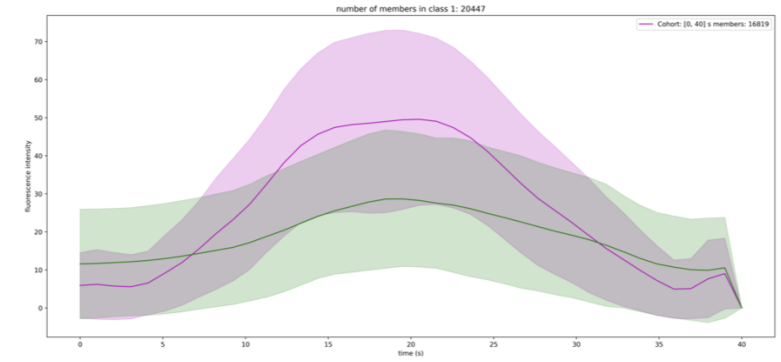
For the maximum intensity alignment, a similar approach is used. First, the index of the peak intensity (DNM2 in our case) is found. Then, each intensity trace is padded with zeros before and after the beginning and end of each trace. Zeros were chosen to mute the effects of outlier events that extend far past the DNM2 peak. For each trace, the number of zeros before and after was calculated using the difference between the maximum of the number of measured values on each end for all tracks and the number in the present intensity trace. This way, all intensity traces within a cohort are of the same length and their peaks are aligned. The intensity traces are then interpolated to span a time domain equal to the ceiling of the cohort. Then, intensities are averaged and plotted.

By first looking at cmeAnalysis+/- events on a cohort-by-cohort basis (Figure 1.11), it is clear that cmeAnalysis- events have little to no DNM2 detections above background levels. However, the variability in phenotypes associated with cmeAnalysis+ group is not clear. cmeAnalysisDNM2+ events show a gradual DNM2 recruitment that peaks near the maximum AP2 intensity as expected for CCPs. Cohort plots of cmeAnalysisDNM2- events show low levels of AP2 recruitment with little to no DNM2 signal above background.

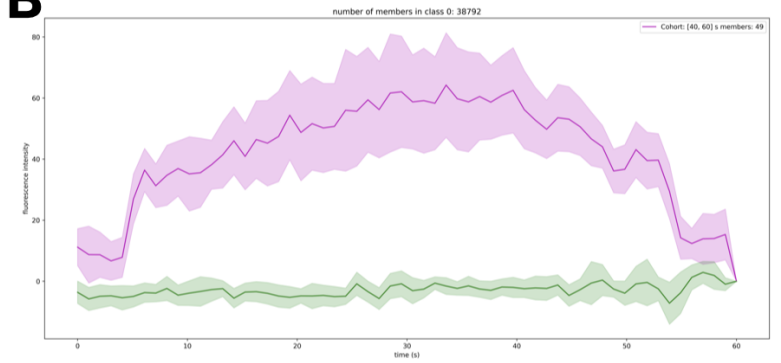
Figure 1.11: aligned intensity traces per lifetime cohort of cmeAnalysis binary categorized tracks based on DNM2 presence

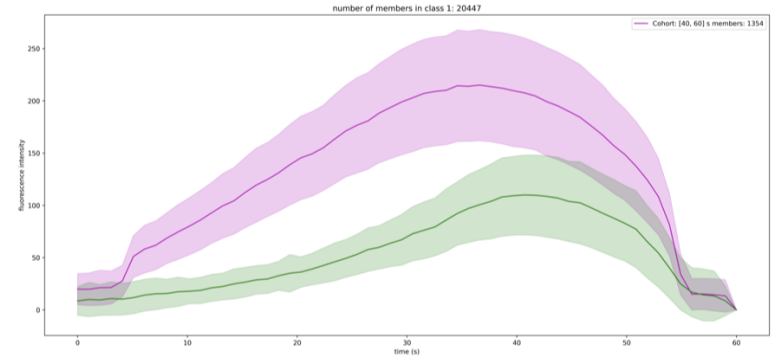
A


number of members in class 0: 38792
Cohort: [0, 40] s members: 38726
fluorescence intensity
time (s)

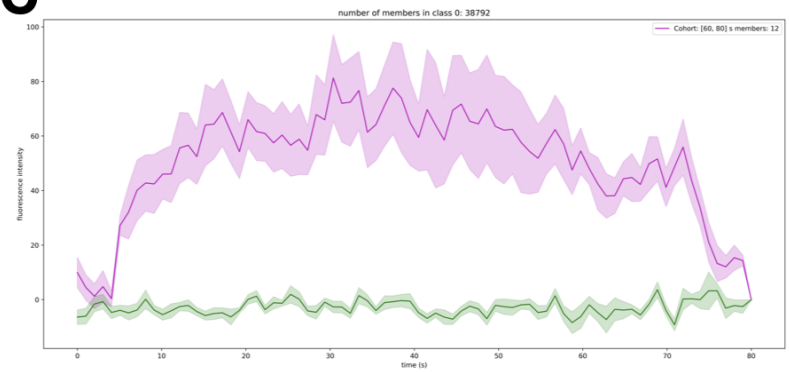

number of members in class 1: 20447
Cohort: [0, 40] s members: 16819
fluorescence intensity
time (s)

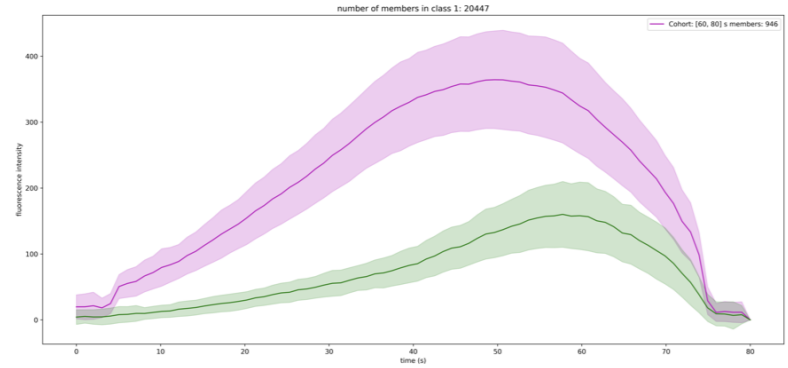
B


number of members in class 0: 38792
Cohort: [40, 60] s members: 49
fluorescence intensity
time (s)

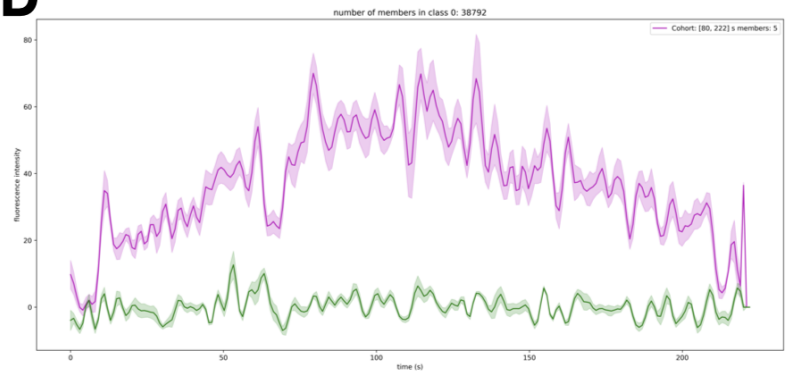

number of members in class 1: 20447
Cohort: [40, 60] s members: 1354
fluorescence intensity
time (s)

C


number of members in class 0: 38792
Cohort: [60, 80] s members: 12
fluorescence intensity
time (s)


number of members in class 1: 20447
Cohort: [60, 80] s members: 946
fluorescence intensity
time (s)

D


number of members in class 0: 38792
Cohort: [80, 222] s members: 5
fluorescence intensity
time (s)

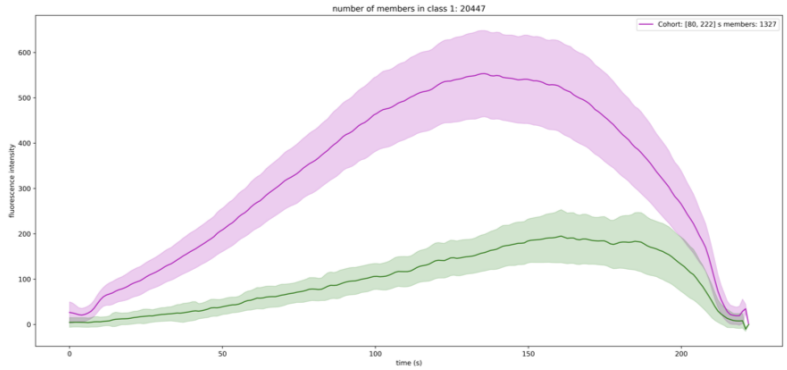

number of members in class 1: 20447
Cohort: [80, 222] s members: 1327
fluorescence intensity
time (s)

Figure 1.11 legend: per A-D, top panel corresponds to cmeAnalysis- and bottom panel corresponds to cmeAnalysis+. Cohorts are [0-40], [40-60], [60-80], and [80-222] seconds. Intensity averages were generated using an interpolation method.

However, when events less than 40 seconds were analyzed broken down by PC-clusters, it was clear that clusters 2 and 4 had DNM2 detections just like cluster 0/DNM2+ events, but the levels were lower (Figure 1.12). Our DNM2+ cluster automatically removes these weakly-recruiting DNM2 events. 71% of cmeAnalysisDNM2+ events that fell into the four DNM2- GMM clusters showed similar average phenotypes: short-lived AP2 recruitment with little to no DNM2 recruitment. These results indicate detection of many short-lived visitor or abortive vesicles to the TIRF field that are within the detection sensitivity of cmeAnalysis, but which must be removed so downstream analysis is limited to authentic CME events.

Figure 1.12: cohort plots per cluster of events with AP2 lifetimes up to 40 seconds with fixed intensity axes

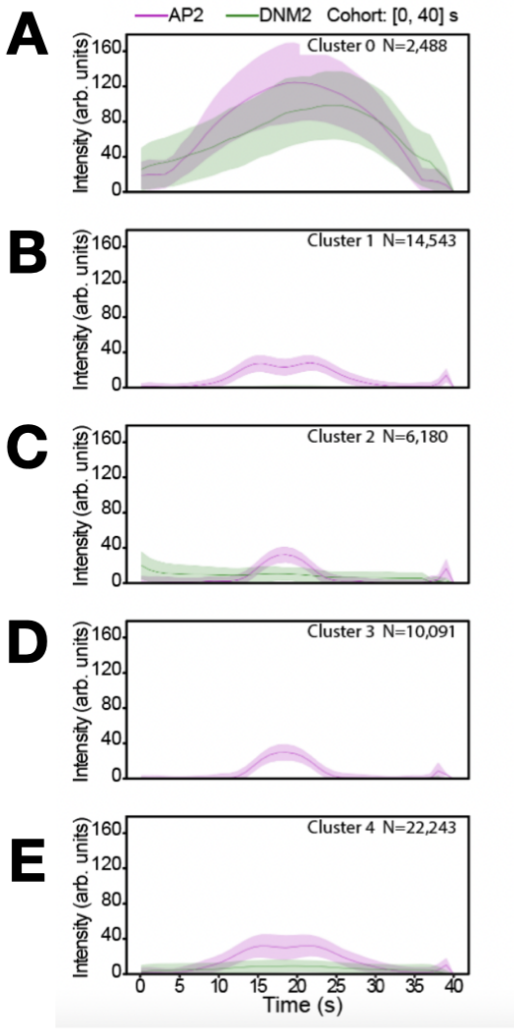
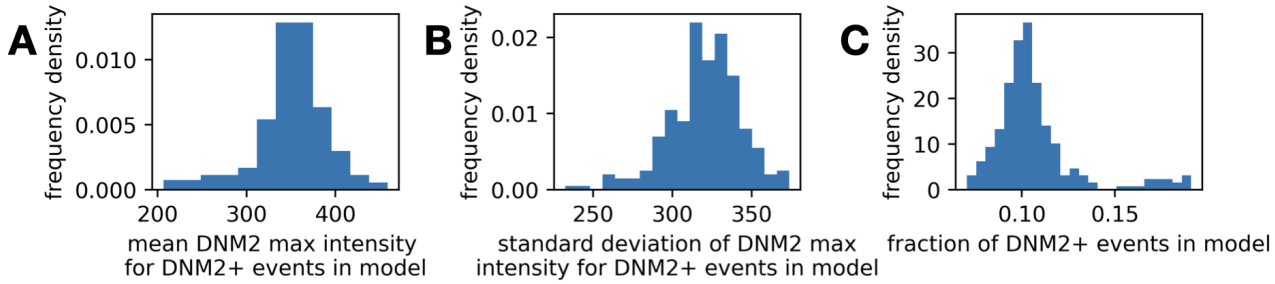


Figure 1.12 legend: A-E: events in clusters 0-4 aligned via the interpolation scheme.

Training dataset influence on selection of DNM2+ events

To understand the influence of training datasets on the selection of cluster 0/DNM2+ events in the aggregate dataset model, I developed a sampling method outlined here: First, every combination of training datasets (8 FOVs) was sampled to generate a candidate selection model with the same PCA and GMM approach previously described. N^2-1 models or 255 models were uniquely generated. Then, every dataset was queried against the candidate selection model and DNM2+ events were identified based on the cluster with the strongest DNM2 signal. Statistics describing the mean DNM2 intensity from each models' DNM2+ cluster are outlined in Figure 1.13 A and B. Roughly ~10% of events from each model were considered DNM2+ (Figure 1.13 C), good agreement with the aggregate model. When viewing the fraction of events that are DNM2+ when a single dataset is used for training (Figure 1.13D), it is clear that DNM2+ selection is influenced more for some datasets than others when compared to the aggregate model's DNM2+ event selection. For instance, experiment #2 (green) has a similar fraction of events selected as DNM2+ regardless of whether 1 (self) or 2-8 datasets are used for training. However, experiment #6 seems differently influenced and has fewer events selected as DNM2+ when training datasets other than itself are used. When comparing the distributions of the fraction of DNM2+ events per experiment across all possible models via a Tukey's pairwise test, we found that most distributions had significantly different means (Figure 1.13 E). From these data, it is clear that which events from each dataset are assigned DNM2+ in the aggregate model is influenced by selection of training datasets.

Figure 1.13: influence of training dataset selection on DNM2-enriched cluster selection



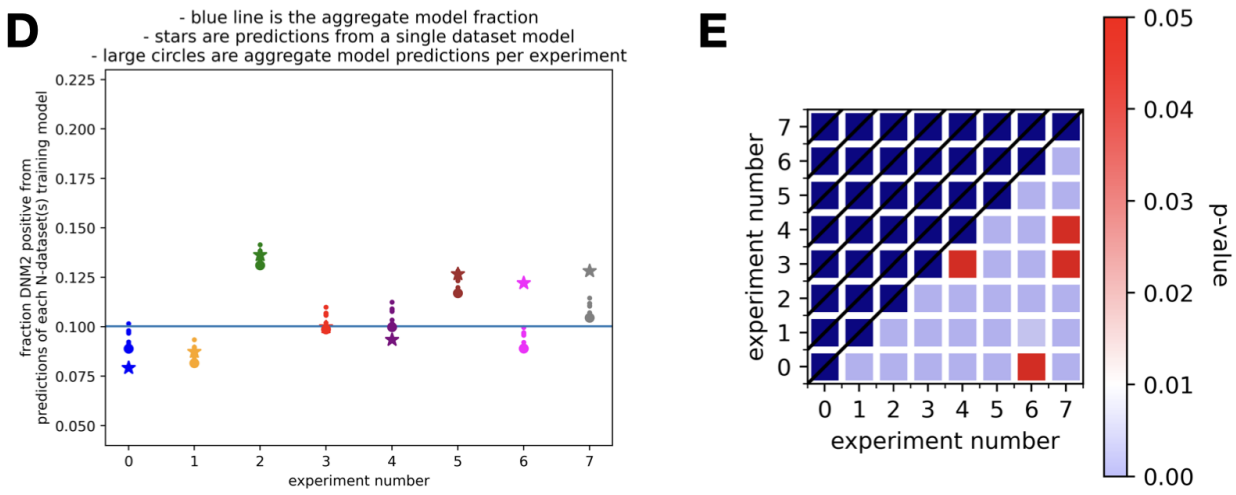


Figure 1.13 legend: A: mean of max DNM2 (per track) intensity histogram from DNM2+ cluster events across all models, B: standard deviation of max DNM2 (per track) intensity histogram from DNM2+ cluster events across all models, C: fraction of events in DNM2+ cluster across all models, D: fraction of events that are DNM2+ per experiment comparing models that have 1 (self), N (2-7) and 8 (aggregate) training datasets, E: p-values of Tukey's post-hoc HSD test of comparisons of % DNM2+ per experiment across all 255 model combinations

To understand the limitations of the aggregate model, we asked what percentage of the possible 255 training models was assigned to each event at DNM2+. Overlaying this with the boundaries of the PC axes highlighted the observation that any disputed event lies between clusters 0 and 4 (Figure 1.14). Indeed, when we looked manually at events on the fringe of clusters 0 and 4, they were often difficult to tell apart. These events were typically "non-peaking" in DNM2 assembly (Figure 1.15 A), indicating a possible imaging artifact from insufficiently fast sampling, or that the biology of these AP2/clathrin events is different in that they do not use DNM2 to mediate scission. These data show us that the events we cared to analyze with AP2 and strong DNM2 enrichment are not heavily influenced by the choice of training data.

Figure 1.14: non-fringe DNM2+ cluster events are universally agreed upon regardless of training datasets

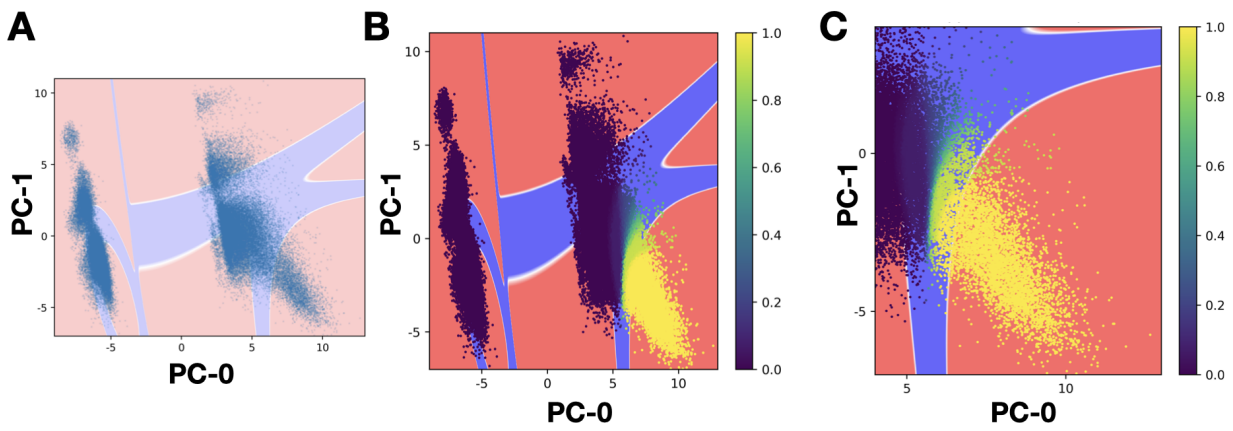


Figure 1.14 legend: A: events from aggregate model in PC space. Background (red/white/blue) represents simulated data points on a grid sampled in space to show posterior probabilities of 95% (white), greater than 95% (red) or less than 95% (blue). 95% was picked arbitrarily. B + C (zoom of B): the disputed events (<95% agreement on DNM2+ from all models) are on the fringe of clusters 0 and 4. Colorbar for B and C corresponds to the percentage of 255 candidate models that consider each event DNM2+.

Figure 1.15: examples of intensity traces from non-peaked, single-peaked, and multi-peaked DNM2 assembly

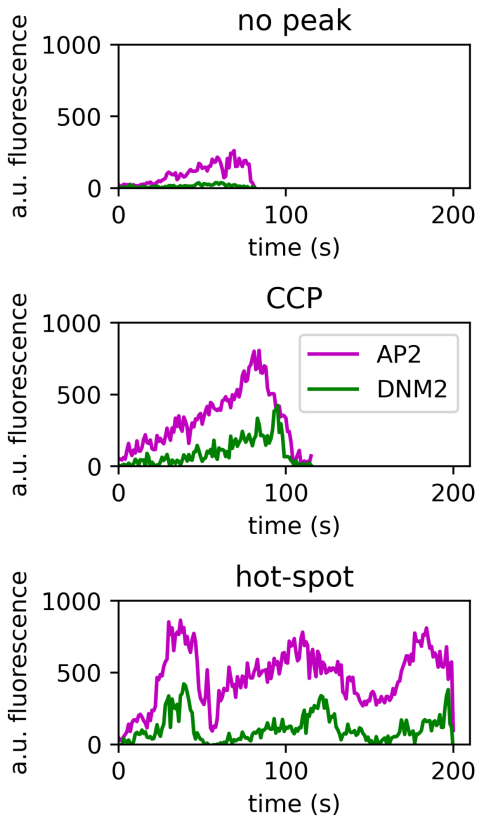


Figure 1.15 legend: Fringe events between clusters 0 and 4 typically had weak DNM2 assembly (A), whereas non-fringe cluster 0 events have one or more DNM2 bursts (B and C).

The identity of the other clusters is not known because we only have information about these events as they transit through the TIRF field, our sensitivity is unknown, and the markers have limitations regarding the events' fate (e.g., whether they recruit early endosomal markers). However, because they generally diffuse rapidly compared to events in cluster 0, are shorter lived, and are less bright, they are likely either abortive CCPs or visitors to the TIRF field (Figure 1.16). Future work with more markers, faster imaging, and defined sensitivities may be useful for discovering the biological origin of these other clusters. The selected DNM2+ events consist of long lived (>~20 seconds), DNM2 recruiting, relatively immobile AP2 structures. Manual verification of events in DNM2- clusters showed the expected phenotype of visitor or abortive vesicles: rapid AP2 appearance that quickly peaks and quickly disappears on time scales less than the minimal expected lifetime of CME events (at least 15 seconds). These intensity characteristics are the expected result of non-plasma membrane bound events that more freely diffuse than their authentic CME counterpart events which nucleate, grow, and pinch off at the plasma membrane.

Figure 1.16: ECDFs of feature differences between DNM2+ and remaining clusters

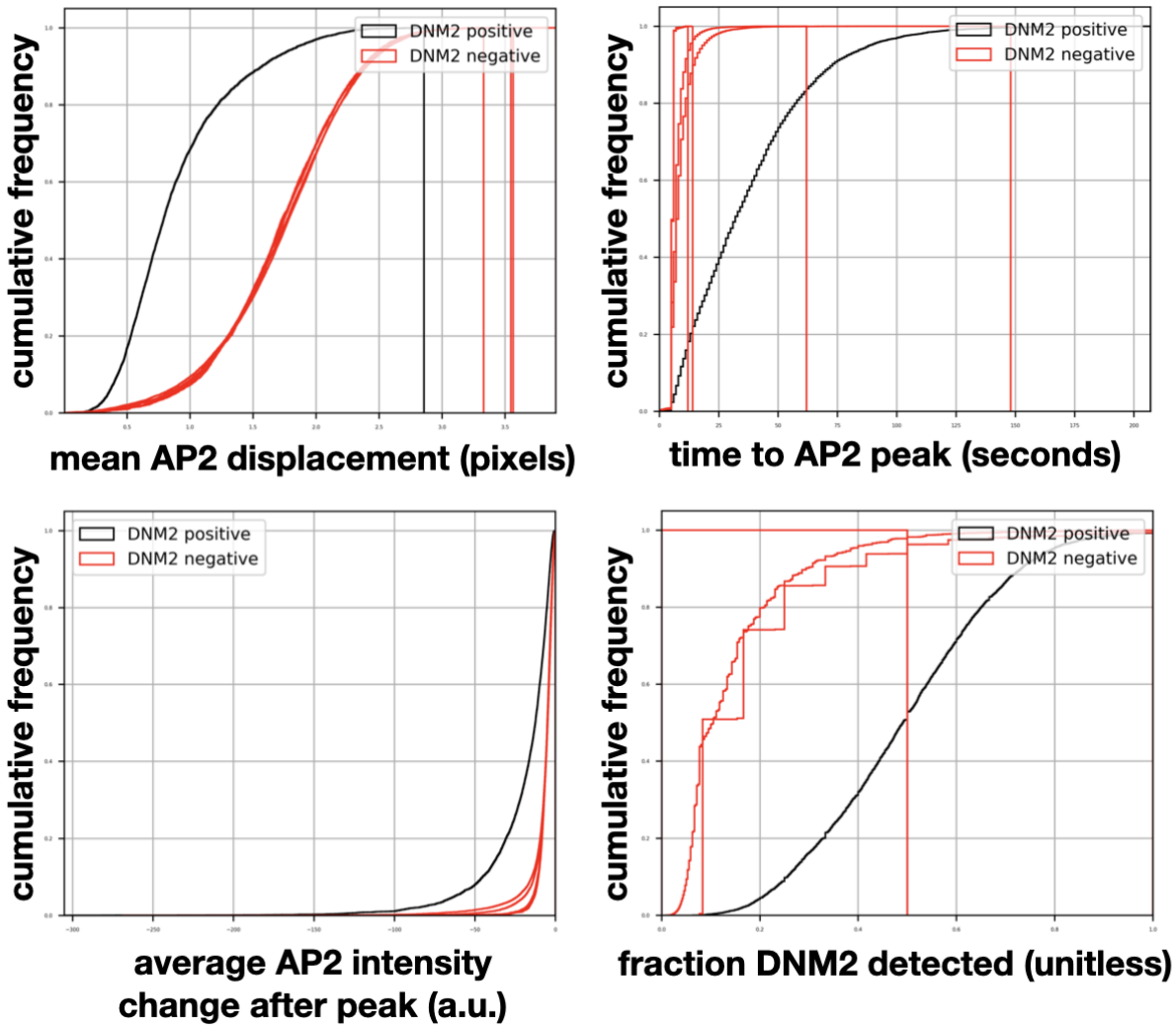


Figure 1.16 legend: A: mean AP2 displacement, B: time to AP2 peak, C: average AP2 intensity change after peak, D: fraction of AP2 track with significant DNM2 detections

DNM2 peak selection modeling

CME is a multi-step process that requires the continual recruitment of proteins from several different functional modules that contribute to the formation of a vesicle. Modeling the kinetics of a multi-step maturation process indicated that the expected lifetimes of clathrin-coated pits (CCPs) should follow a Rayleigh distribution[13]. Rayleigh processes describe events with an increasing rate of failure as lifetimes increase. The lifetime distribution of all events following the merging of all DNM2- positive events marked by cmeAnalysis resulted in an exponential-like distribution. Furthermore, an attempted fit to a Rayleigh distribution resulted in a poor fit, which was confirmed by a Kolmogorov-Smirnov goodness-of-fit test with a p-value of 0 under the null hypothesis that the data are sampled from identical distributions (Figure 1.17)

Figure 1.17: AP2 lifetime distribution of DNM2-enriched events from cmeAnalysis binary classification

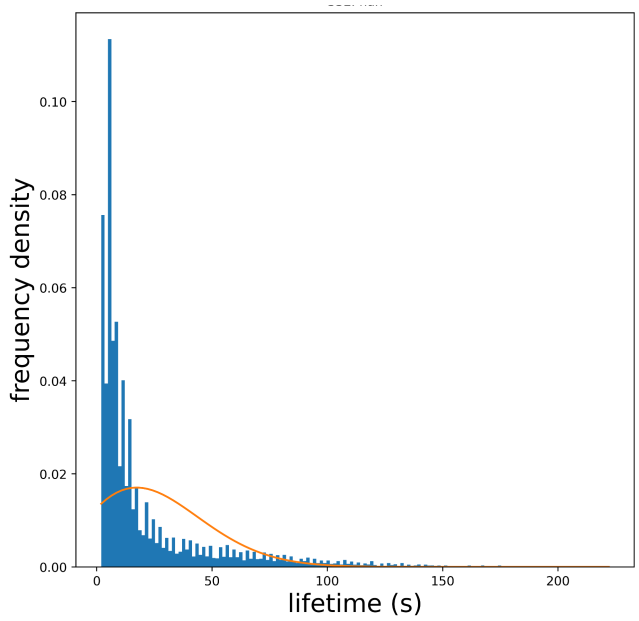


Figure 1.17 legend: histogram of cmeAnalysis+ events (blue) with poor Rayleigh distribution fit (orange)

So far, we isolated DNM2 events that are relatively immobile and have strong recruitment of AP2 and DNM2. These DNM2+ events are not completely designated as CCPs since they have variable numbers of visible DNM2 peaks (Figure 1.15). Also, some DNM2+ events weakly recruit DNM2 for long periods of time without forming a sharp burst. Therefore, further refining events into “authentic CCPs,” which are marked by single DNM2 bursts, is imperative. Multi-burst events, or “hot-spots” are known to be phenotypically distinct from de novo nucleation of AP2 followed by vesicle scission (reference). Also, due to the temporal resolution of our imaging (1 second), we are unable to capture and identify events taking place on time scales of 2 seconds or less, so some events may not exhibit a clear DNM2 peak, making them difficult to classify as DNM2-mediated scission events. Furthermore, there may be events that do not primarily rely on DNM2-mediated scission.

This work makes, to our knowledge, the first attempt to classify tracked CME events by the number of DNM2 bursts they contain. Our strategy was to identify the characteristics of a single DNM2 peak and then count how many DNM2 peaks, if any, show up in an event. Our approach to finding DNM2 peak characteristics was inspired by the assertion that authentic CCP events have a Rayleigh distribution of lifetimes. Therefore, we turned to a parameter sweep where we tested which minimum peak-describing requirements yielded single-DNM2-peak events with lifetimes closely resembling a Rayleigh process.

We first showed that our DNM2+ events appear more similar by eye to a Rayleigh distribution, but still fit poorly by three standard goodness-of-fit metrics (Figure 1.18 A). In particular, there is a large contribution of very short lived, DNM2-rich events that

creates a distinct peak in the lifetime distribution near 10-15 seconds. The first distribution comparison was made with a two- way Kolmogorov-Smirnov (KS) test. However, it is important to note that since the raw distribution is forced to curve-fit a Rayleigh distribution, this test's results can be misleading when compared to the empirical data. So, we supplemented the KS result with a Chi-squared goodness-of-fit test and sum-of-squared errors (SSE) comparison. The Chi-squared test was executed as follows: the cumulative observed frequency was estimated by taking the lifetime percentiles at 2% intervals and binning a histogram of lifetimes at these lifetime percentiles. Then, an expected cumulative frequency distribution was measured using Rayleigh-fit parameters at the lifetime percentile thresholds. Next, for each percentile bin, the fraction of all events in each bin was calculated and used to generate a discretized, cumulative frequency array. Finally, these cumulative frequency distributions were compared with a one-way, Chi-squared test. The SSE is measured in the following way: observed frequencies are computed from a normalized-to-one histogram which is sampled at 2% percentile intervals. Next, a probability density function is estimated from the Rayleigh fit parameters and sampled at the bin edges from the histogram. Finally, the difference between the observed frequencies and estimated frequencies are squared and then summed.

Figure 1.18: AP2 lifetime distribution of events from PCA-based DNM2-positive cluster, and DNM2 frequency modes

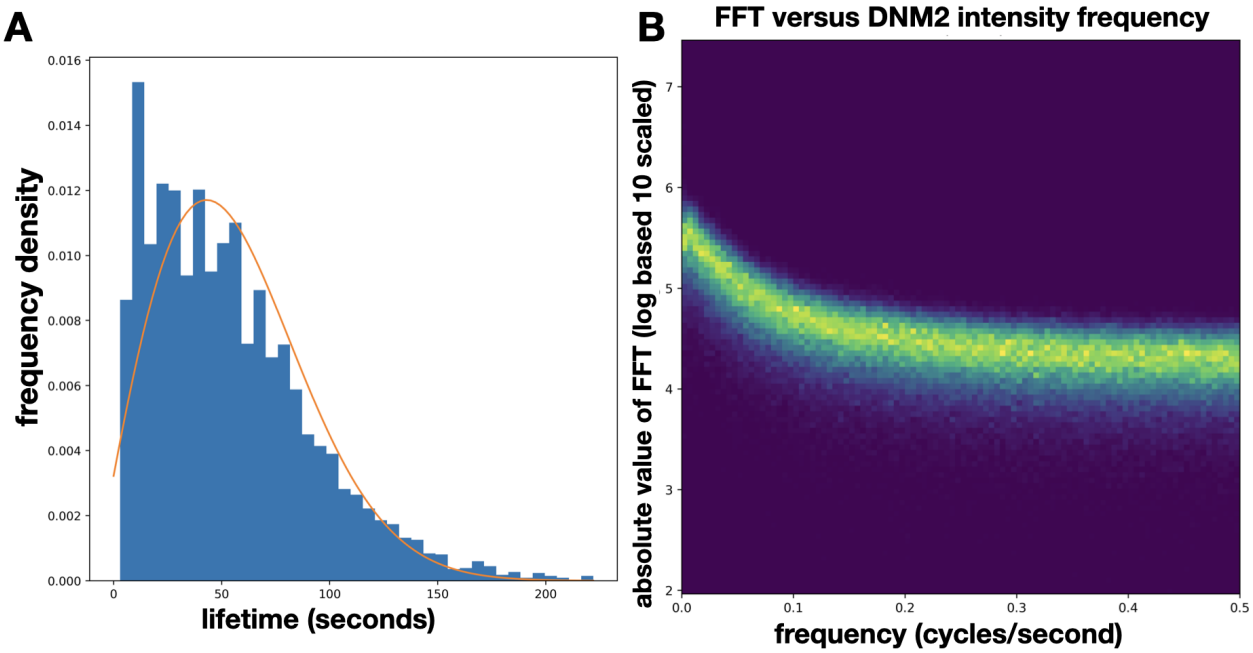


Figure 1.18 legend: A: histogram of AP2 lifetimes from DNM2 + cluster which appears less exponentially distributed but still poorly fits a Rayleigh distribution. p-value of Kolmogorov-Smirnov test: 1.32e-9, p-value of Chi-squared test: 0.0, SSE: 0.00034, B: 2D histogram of the absolute value of Fast Fourier Transform (log-scaled) versus frequency.

In order to aid in filtering out slow oscillations in intensity representing DNM2 bursts, a signal filter was applied to DNM2 intensities to remove high-frequency noise. We applied a Fast Fourier Transform (FFT) to visualize the contributions of frequencies of all DNM2 intensity traces from DNM2+ events stitched together (Figure 1.18 B). The available frequencies are hard-capped by the Nyquist frequency which is half the sampling frequency or 2 seconds per frame. The Nyquist frequency describes the fastest oscillation that can be used in a Fourier reconstruction of our DNM2 waveforms, since we are imaging at twice its rate or 1 second per frame. The two-dimensional histogram of FFT intensities versus sample frequencies shows two features. First, there is a near-homogenous contribution of high-frequency signals present in DNM2 intensity traces that are quicker than 0.2 cycles per second. Below this frequency, there is a broad distribution of slower oscillations that are up to an order of magnitude more intense than the high-frequency signals. This result indicates the existence of a non-uniform range of slow frequencies that correspond to possible DNM2 peaks. The 0.2 cycles per second elbow served as the cut-off frequency used on all DNM2 intensities via a low-pass, fourth-order Butterworth filter.

Previous work using molecule counting has established a basis for the minimum number of DNM2 molecules necessary for vesicle scission. However, since our imaging was not calibrated for single-molecule sensitivity, we needed to discover a proxy for sufficiently-bright peaks that correspond to the minimum DNM2 requirement for vesicle scission. Additionally, quickly-appearing and bright signals can appear in intensity traces as a result of unbound vesicles appearing near a tracked event. Therefore, it is important to establish the requirements for minimum peak widths that correspond to *de facto* DNM2 recruitment and disassembly. Finally, since multiple peaks can appear in a single event, it is important to establish the minimum time requirement between peaks. This serves to eliminate false-positive peaks that appear too close to prominent peaks which can result in an event being misclassified as multi-peaked.

Our algorithm's outline is as follows: (1) select from a combination of three peak-requirement parameters (minimum height, minimum width, and distance between peaks), (2) find which events have filtered DNM2 intensities with one peak, and (3) perform a statistical comparison between the lifetimes of single-peaked events and a Rayleigh distribution fit from Maximum Likelihood Estimation. The combinations of peak requirements span a meshed grid housing all combinations of the three parameters across selected ranges. The minimum height was chosen to be 50 to 300 a.u.'s in 25 a.u. intervals since the DNM2+ cohort plot peaks were within this range. The minimum peak width was chosen as 1 to 10 seconds to cover the sampling frequency (inaccessible frequency modes below Nyquist) to time scales nearing the minimum recorded CME lifetimes (15 seconds). The minimum distance between peaks was chosen as 1 to 20 seconds to cover the length of lifetimes around the minimum recorded CME lifetimes. We used the Kilmogorov-Smirnov (KS) p-value maximum to select the best-fit parameters and validated this goodness-of-fit with the Chi-squared p-value and SSE measurement. The KS test implementation was working under the null hypothesis that the distributions are identically distributed.

The optimal parameters to describe a DNM2 peak were a height of 125 a.u.'s, 5 second width, and 17 seconds between peaks, which were found in 42% of DNM2+ events. The p-value for the KS test was 0.71, the p-value for the Chi-squared test was 0.97, and the SSE was 0.00027 between the observed and estimated probability density function (Figure 1.19). The minimum peak height is in close agreement with the peak amplitude found in interpolation-based cohort plots across all lifetime bins. However, this interpolation method distorts the maximum of each event since they do not necessarily occur at the same index in the interpolated time domain. The minimum peak width is, to our knowledge, not a parameter that has been measured in the field. However, the minimum CME lifetimes referred to were imaged with single-molecule sensitivity, making their 15 second minimum a more accurate lifetime since the first and last clathrin molecules between initiation and vesicle uncoating were captured[9].

Figure 1.19: successful Rayleigh-distribution fit of AP2 lifetimes from single-peaked DNM2 events

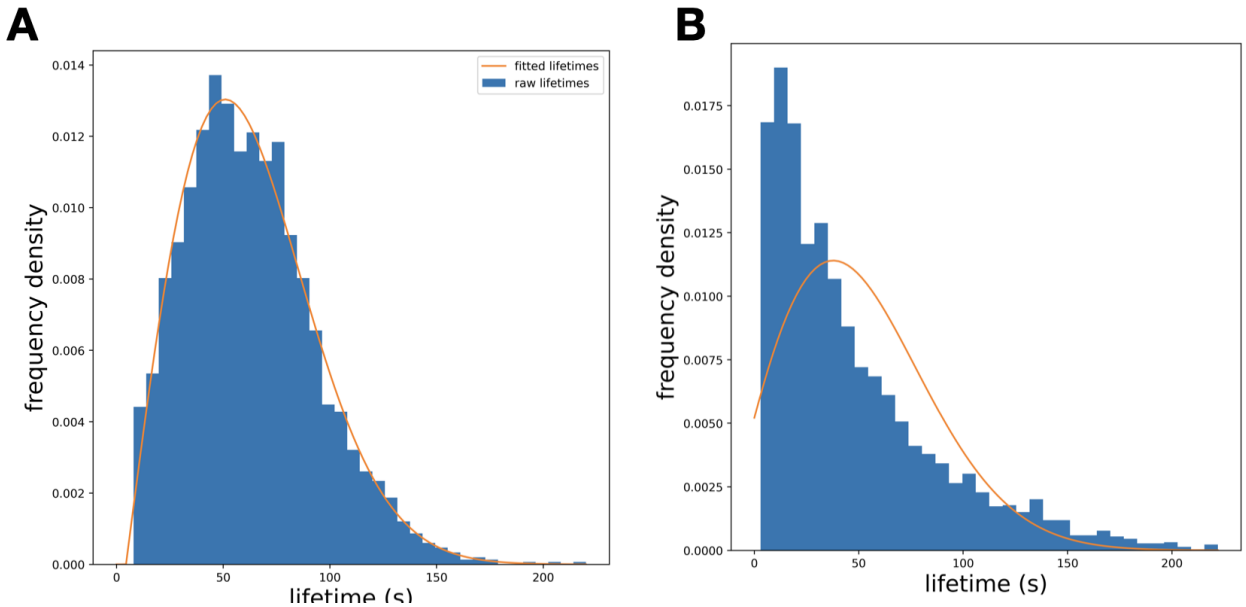


Figure 1.19 legend: left: AP2 lifetimes of single-peaked DNM2 events (1 model, highest KS-test value) in blue and Rayleigh distribution fit in orange, p-value of Kolmogorov-Smirnov test: 0.71, p-value of Chi-squared test: 0.97, SSE: 0.00027, right: lifetimes of non-peaked and multi-peaked events remaining from cluster 0/DNM2+ events in blue and Rayleigh distribution fit in orange, p-value of Kolmogorov-Smirnov test: 1.72e-59.

When comparing all "good" peak-finding models (Figure 1.20), i.e., those models with KS-test p-values greater than 0.01, we found that DNM2-peaking events were typically not situated close to the boundary of clusters 0 and 4, whereas non-peaking events were grouped tightly near the fringe of the two clusters. As the posterior probability of being in cluster 0 decreases, there is a decrease in the number of detectable peaks, Figure 1.21 B and C. In fact, there appears to be a positive correlation with PC-0 and the number of peaks, as well as a negative correlation with PC-1 and the number of peaks (Figure 1.22) Further work could be done to generate bootstrapped models that

consider the selection of CCPs not based on a single model's peak-finding constraints, but compare multiple models to rule out agreed upon non-peaking events that could be the result of non-DNM2-mediated scission.

Figure 1.20: Choices of events with N-peaks for top-performing DNM2 peak selection models

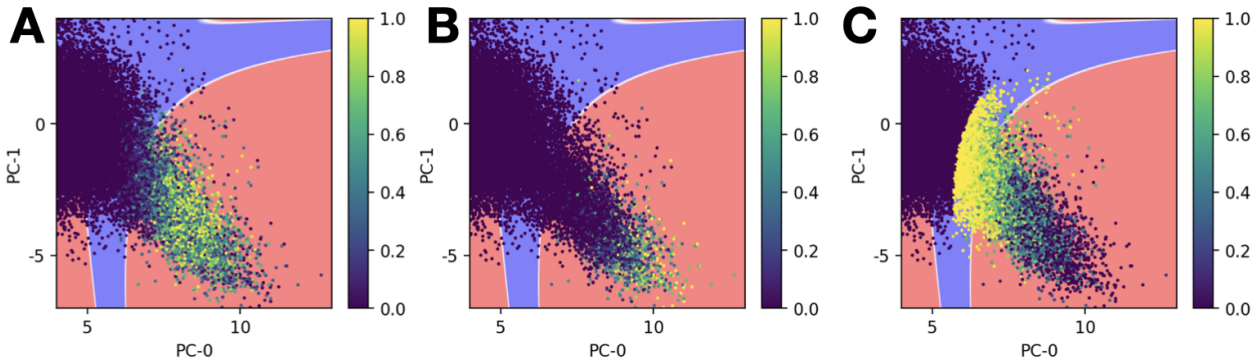


Figure 1.20 legend: Only considering models for which the single-peaked events had a KS-test p-value greater than 0.01 and zooming into cluster 0 with 95% posterior probabilities outlined in white from simulated GMM inference data points. Red: posterior probability is greater than 95% to nearest cluster center; blue: posterior probability is less than 95% to nearest cluster center. Color-bar: percentage of 255 candidate models that considered individual data points A: CCP, B: multi-peaked, C: non-peaked.

Figure 1.21: DNM2 peak relationships with Principal Components

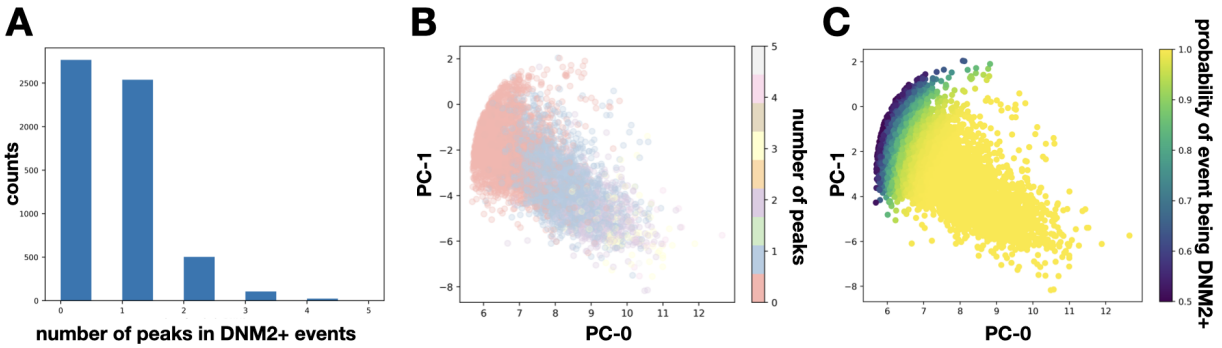


Figure 1.21 legend: A: raw counts for DNM2 peak counts from cluster 0/DNM2+ events, B: number of peaks in events from DNM2+ cluster overlaid on scatterplot of principal component coordinates, C: posterior probability for hard clustering from GMM of real data points in scatter coordinates of principal components. Note the drop-off in probability (yellow to purple) as PC-1 gets better and PC-0 gets smaller is coincident with predicted number of peaks (B - pink).

Figure 1.22: DNM2 peak relationships with Principal Components

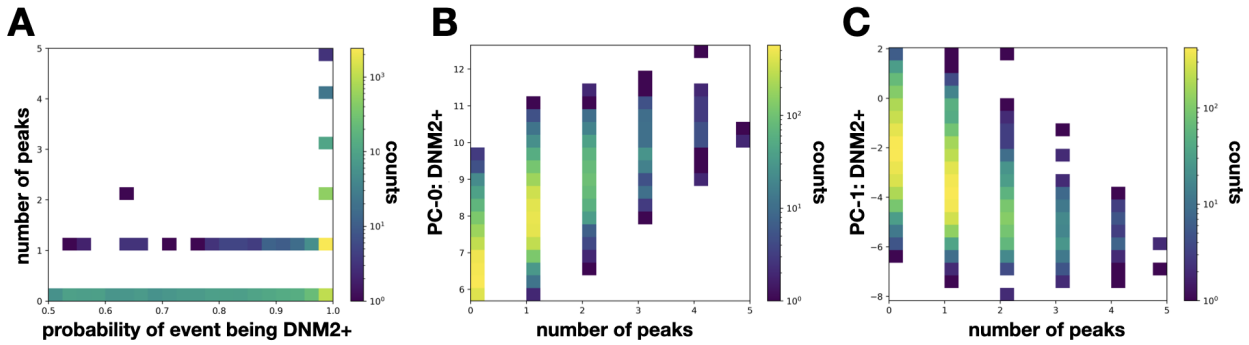


Figure 1.22 legend: 2-D count histograms. A: posterior probability of events in DNM2+ cluster against (integer) DNM2 peak counts, B: PC-0 versus number of peaks, C: PC-1 versus number of peaks.

The single-peaked DNM2+ events will hereon be referred to as CCPs to differentiate them from multi-peaked hot-spots and non-peaked events without detectable DNM2 bursts. These model-mined parameters are a significant step toward using consistent heuristics in selecting relevant events in studies of CME or dynamics of other organelles. Importantly, when making comparisons across various experimental conditions, it is imperative that criteria for the properties of wild-type assembly dynamics to use as a baseline for inferring the conditions found in new datasets are established. While this approach does not eliminate the need for, or underscore the importance of, molecule counting, it does offer a possible workaround for determining successful scission based on reverse modeling expected lifetime distributions. Further studies will be necessary to validate the results of the selected DNM2 peak characteristics. These results, when combined with future molecule-counting studies, could allow us to reach a consensus on how many DNM2 molecules are required for vesicle scission. These modeling efforts aim to bridge the gap between the need for robust functional read-outs of vesicle scission and the feasibility of such necessary experiments.

Now that we have CCP requirements, we find that of the 59,239 valid tracks used to build the model, 10% of all events are DNM2+ and 4% of all events are CCPs. The remainder of the DNM2+ events had either no peaks (46%) or two or more peaks (10.60%). Figure 1.23 summarizes the classification schemes described so far: PCA-based clustering and inverse modeling of DNM2 peaks using target AP2 lifetime distributions.

Figure 1.23: PCA clusters and visual overlaps with cmeAnalysis binary classification

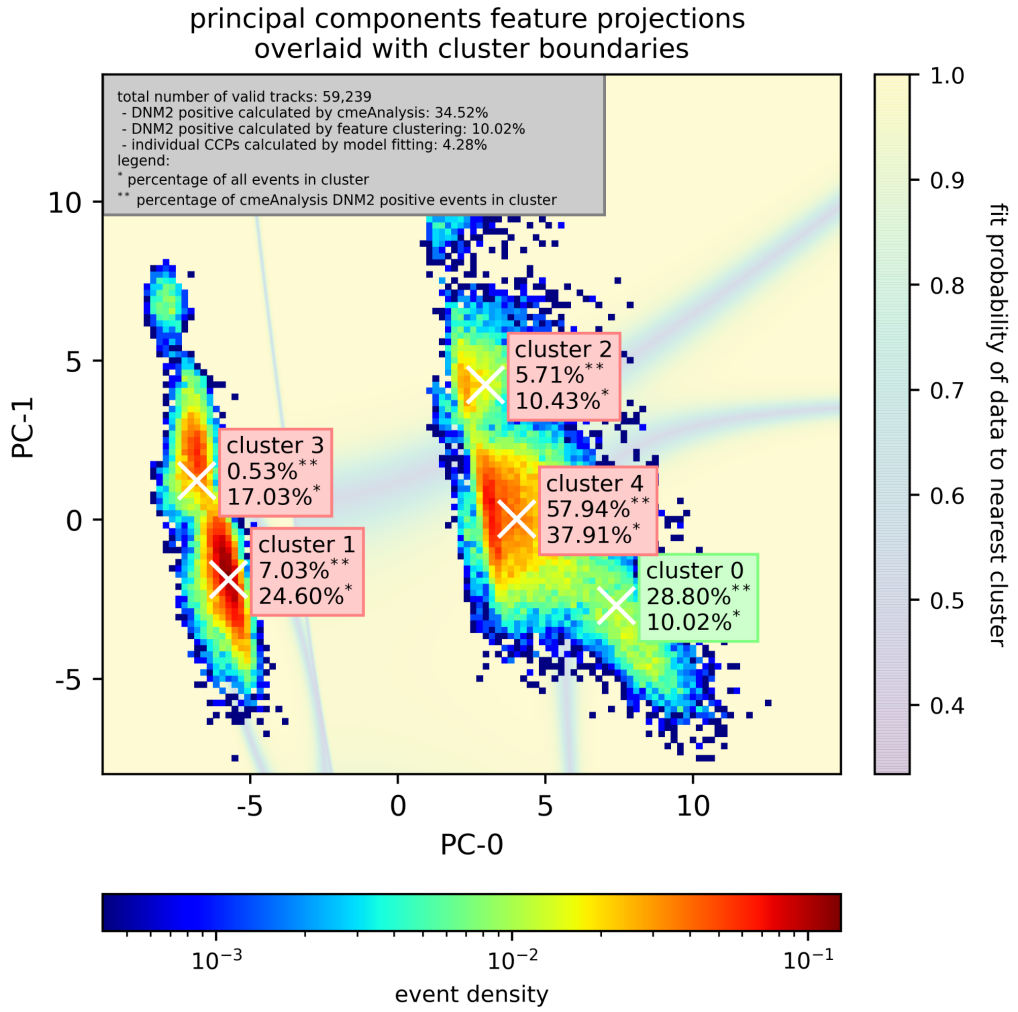


Figure 1.23 legend: summary of cluster overlaps with cmeAnalysis+/- and CCP selection

Binned and aligned-to-peak- DNM2 CCPs showed the expected phenotype of AP2/DNM2 dynamics: gradual AP2 assembly accompanied by weak DNM2 recruitment that bursts near the peak of AP2 recruitment. To verify that DNM2- clusters did not contain sufficient DNM2 peaks to mediate vesicle scission, we searched for DNM2 peaks across the other GMM component events. Less than one percent of “cluster 4” events contained DNM2 peaks and they were entirely absent in other clusters. As a reminder, “cluster 4” was labeled as putative “DNM2-carrying visitors”, so the presence of occasional bright DNM2 signals was not entirely surprising.

Table 1.3: fractions of events with N-peaks between PCA clusters

number of peaks	fraction of events in: cluster 0	cluster 1	cluster 2	cluster 3	cluster 4

	(DNM2+)				
0	0.46	1.00	0.99	1.0	0.99
1	0.42	0.00	0.00	0.00	0.01
2+	0.11	0.00	0.00	0.00	0.00

We verified the average assembly dynamics of non-peaked DNM2+ events and found a surprising result: the lifetime binned and aligned-to-DNM2-peak events showed characteristic AP2/DNM2 recruitment (Figure 1.22). However, it is important to note when comparing the longest-lived non-peaked cohort (>=80 second AP2 lifetime) and the shortest single-DNM2-peaked cohort (<40 second AP2 lifetime), that the two had a similar peak AP2 brightness but the single-peaked events had nearly double the DNM2 brightness.

Figure 1.24: the strongest non-peak DNM2 tracks have weaker DNM2 recruitment than the weakest CCPs

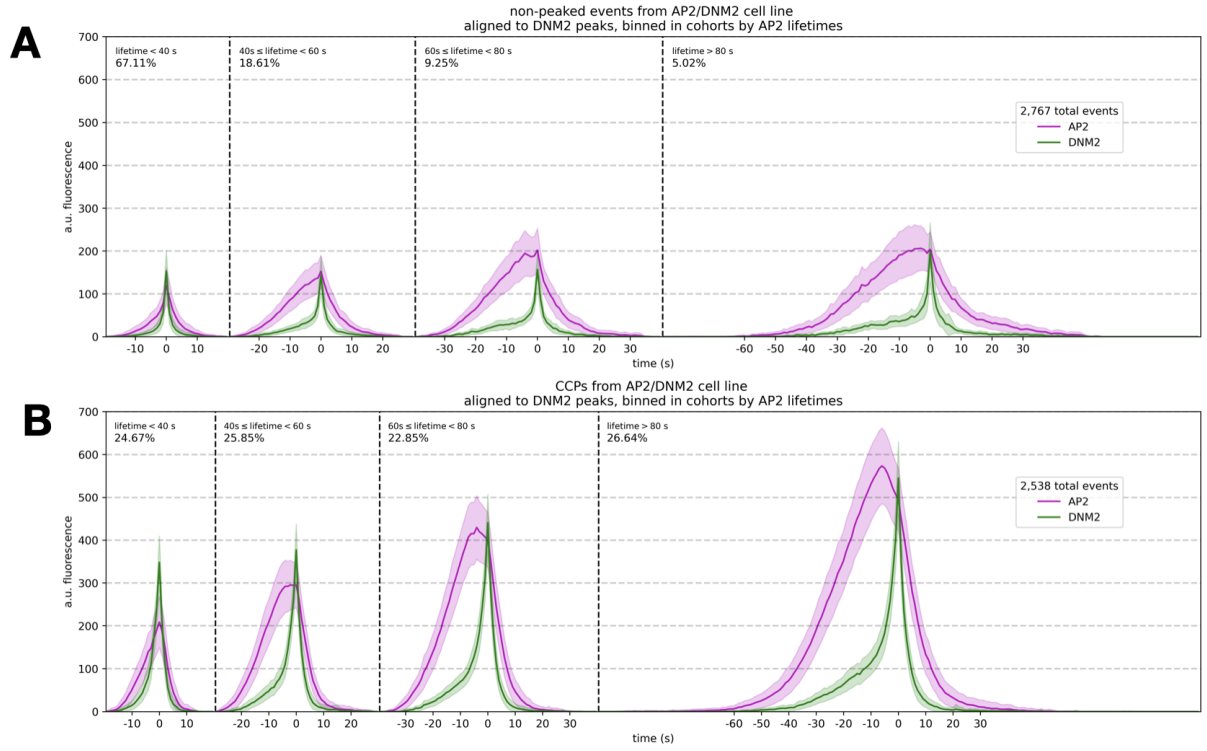


Figure 1.24: aligned intensity traces of AP2 and DNM2 split between CCPs (bottom) and non-peaking events (top) from DNM2+ cluster. 4 cohorts: lifetime<40, 40<=lifetime<60, 60<=lifetime<80, lifetime>80 seconds. Threshold for minimum DNM2 intensity for peak was 125 a.u.'s.

Discussion

In this chapter, I present a generalizable method of finding principal modes in particle tracking data. These approaches could generalize to any data for which there are tracked objects with statistical measures of their detection confidences. When multiple objects are tracked in different channels, the different combinations of their co-localizations can be projected via interpretable analysis (PCA), then reconstructed to useful coordinates (e.g., traces of intensity through time).

The standards for selecting events relevant to a study of CME dynamics often involve arbitrary event selection and filtering criteria with few easy-to-interpret standards. Cherry-picking relevant data is an easy way to fall into traps of misinterpretation and false conclusions. I hope these efforts will move the studies of CME dynamics in a direction where biological insights can be made more clear by removing events that can be classified as noise.

It would be useful to repeat these experiments but image with calibrated single-molecule sensitivity to know the protein copy number that corresponds to the boundaries between clusters. This information could help identify the biological origin of the different non-DNM2 clusters: visitors, abortive events, or different types of faster endocytosis. Additionally, integrating other markers (e.g., endosomal markers such as Rab GTPases) may be useful to label visitors.

Another area for future exploration is bootstrapping the DNM2 peak detection modeling. After all, there were plenty of models for DNM2 peak constraints that provided a well-fit Rayleigh distribution of single-peaked events. When compiling their predictions, it was clear that they disagreed on the CCPs but agreed well on non-peaking events (Figure 1.20 C). Furthermore, those agreed-upon non-peakers were events that had low probability of being DNM2+ (Figure 1.21 C). I think it would be useful to combine this information to be able to categorize non-DNM2-peakers from DNM2-peakers (1+ peaks).

Furthermore, being able to move between cell lines with different markers is a ripe opportunity for this approach of data analysis. For instance, if we could compile requirements for “CCP” (1+ DNM2 bursts) using a cell line that has AP2/DNM2 (and/or possibly another late stage marker (Auxillin/GAK)[9]), we could then use supervised machine learning to predict the identity of events using only the AP2 features. Then we could free up fluorophore channels and change the secondary markers (DNM2/Auxilin) to predict CCPs that recruit downstream markers and instead study another marker (e.g., one or more actin effectors).

Methods

All analysis was performed with Python 3 in Jupyter Lab with routine calculations performed in a Python package “cmeAnalysisPostProcessing”. All code and data for these models can be found at “https://github.com/DrubinBarnes/Jin_Shirazinejad_et_al_branched_actin_manuscript”,

with instructions on how to set up the Conda environment specifying every library version used for analysis: "https://github.com/DrubinBarnes/Jin_Shirazinejad_et_al_branched_actin_manuscript/tree/main/conda_configuration". Standard scientific Python libraries such as numpy, scipy, pandas, seaborn, scikit-image, and scikit-learn were used for all calculations and data visualization.

Chapter 2: Arp2/3 complex dynamics at CCPs in human stem cells

Abstract

Actin assembly facilitates vesicle formation in several trafficking pathways, including clathrin-mediated endocytosis (CME). Interestingly, actin does not assemble at all CME sites in mammalian cells. How actin networks are organized with respect to mammalian CME sites and how assembly forces are harnessed, are not fully understood. Here, branched actin network geometry at CME sites was analyzed using TIRF microscopy and the novel event identification methods discussed in Chapter 1. When endocytic dynamics of unperturbed CME sites are compared, sites with actin assembly show a distinct signature, a delay between completion of coat expansion and vesicle scission, indicating that actin assembly occurs preferentially at stalled CME sites. In addition, the Arp2/3 complex is recruited to one side of CME sites, where it is positioned to stimulate asymmetric actin assembly and force production. We propose that actin assembles preferentially at stalled CME sites where it pulls vesicles into the cell asymmetrically, much as a bottle opener pulls off a bottle cap.

Introduction

Formation of clathrin-coated vesicles requires forces to first bend the membrane into a sphere or tube, and to then break the thin neck that connects the vesicle to the plasma membrane. These forces are generated through the combined actions of proteins that directly bend the membrane and actin filament assembly (Figure 2.1)[20,21,23–25]. Several studies have demonstrated that dependence of CME on actin assembly increases under elevated membrane tension[26–30]. Interestingly, actin does not assemble at all CME sites in mammalian cells, suggesting highly localized differences in requirement for actin assembly, the nature of which is obscure[31,32]. A detailed understanding of how actin forces are harnessed to aid vesicle formation and scission depends on understanding which CME sites assemble actin, where filament assembly occurs around the endocytic membrane, and when. In yeast cells, where turgor pressure is particularly high, super-resolution data suggest that actin assembles symmetrically around CME sites

and indicate that actin regulators including Las17, which is yeast WASP, are present in a ring surrounding the base of the clathrin coat symmetrically[33]. On the other hand, studies on fixed mammalian cells raised the possibility that actin assembly may at least in some cases be initiated asymmetrically at clathrin coats[34,35]. However, methods used for these studies prevented analysis of large numbers of sites, and suffered from possible loss of actin filaments during unroofing and extraction of the cells. Which CME sites assemble actin, and how actin networks are organized with respect to CME sites, has not been determined systematically in a large-scale, unbiased manner, particularly in live mammalian cells. This information is essential to understanding how and why actin assembly forces are harnessed for CME.

Figure 2.1: Overview of CME modules studied in Chapter 2

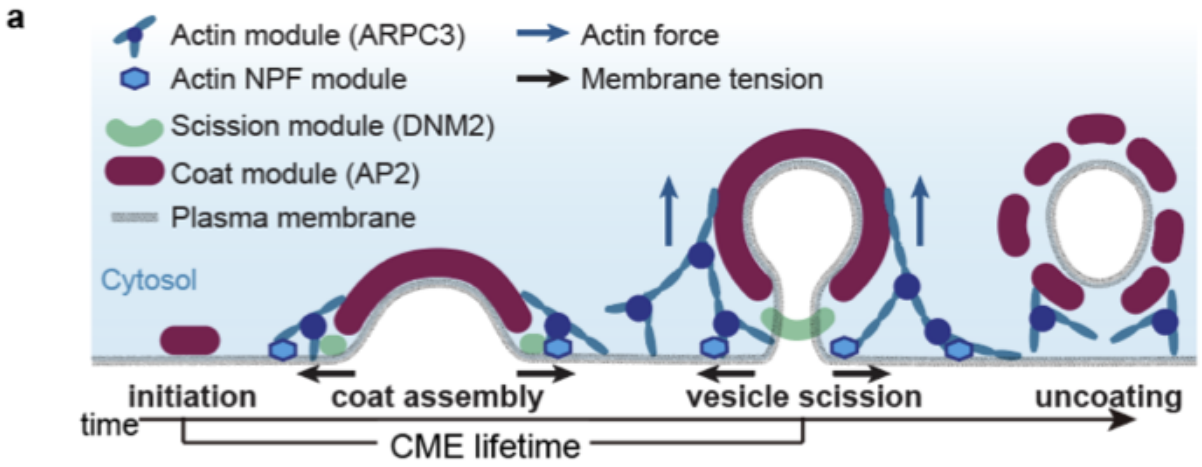


Figure 2.1 legend: a Schematic model of CME. Mammalian CME proteins can be grouped into several modules, including the coat, WASP and Myosin / actin nucleation promoting factor (NPF), actin and scission modules. Actin networks provide pulling forces to invaginate the membrane against membrane tension.

Here, by combining fixed and live-cell imaging of triple-genome-edited, human induced pluripotent stem cells (iPSCs), and newly developed machine-learning-based computational analysis tools, we report that N-WASP and the Arp2/3 complex localize at one side of the coat and neck of invaginating endocytic sites until vesicle scission. Most importantly, by comparing recruitment dynamics of proteins from three distinct endocytic modules for over one thousand unperturbed endocytic events, we found that branched actin assembly occurs predominantly at sites that have stalled between coat expansion and vesicle scission. We propose that these branched actin networks rescue stalled CME.

Results

Actin networks assembled at CME sites remain asymmetric through scission

To analyze the intrinsic recruitment order and timing for three endocytic proteins at CME sites quantitatively and systematically, we developed an automated, high-throughput method to analyze TIRF live-cell imaging data that avoids bias because it does not involve manual selection of CME sites (see Chapter 1). Briefly, AP2 tracks were identified using standard particle-tracking algorithms[13]. Novel filtering methods then extracted DNM2-positive events marked by one or more DNM2 burst. The AP2 and DNM2 tracks were decomposed into dynamic features describing each events' position and brightness. These features were used for clustering via unsupervised machine learning, which enabled grouping of similarly-behaved tracks (Figures 1.9 and 1.12). DNM2-positive events were refined by a detection scheme that determined the number of DNM2 peaks using various characteristics of a single DNM2-peak: the peak height, width, and minimum peak-to-peak distance (Figure 1.15). Events with a single DNM2 peak were analyzed as described below. The method detects low signals from endogenously tagged CME proteins, such as the low-level recruitment of DNM2 at the early CME stages, and accurately reveals the different CME stages (Figure 1.24).

Next, the timing of actin network assembly at CME sites was determined using ARPC3 as a branched actin filament marker by analyzing over one thousand CME events. Although actin appearance early in CME has been reported[34], determining the actin assembly timing can be challenging because it is difficult to distinguish newly assembled branched actin at CME sites from the nearby cortical actin filaments or actin filaments associated with other vesicles or organelles. Also, whether actin functions during the early stage of CME has not yet been shown conclusively due to the potential side effects such as changes in membrane tension caused by actin inhibitors. Our endogenous ARPC3 tagging and large-scale computational analysis approach sidesteps these problems. We classified CME events into three groups: one group without ARPC3 appearance, one group with ARPC3 appearance early in CME, and one group with ARPC3 appearance late in CME (Figure 2.2). We observed that in most of the ARPC3 positive events a sharply increasing ARPC3 signal appears with similar timing to the rapid-recruitment phase of DNM2 concomitant with the U to Ω membrane shape transition (Figure 2.3 a). This timing is consistent with the previously proposed role for actin in membrane invagination, as studies showed that actin inhibitors block the U to Ω endocytic membrane shape transition[26,34]. In some cases we did detect ARPC3 signals at early CME stages (Figure 2.2 b). To test whether random overlap between nearby actin structures and CME sites might be responsible for the apparent early actin recruitment, we generated a randomized data set by pairing ARPC3 images with AP2

and DNM2 images from an unrelated movie (Figure 2.2 a). In this data set, we detected a significantly reduced fraction of ARPC3 positive events. However, early “assembly” of actin was observed in a similar fraction of CME events as in the real data set (Figure 2.2 b). Based on these observations we conclude that the presence of actin early in CME is very likely due to nearby actin structures overlapping with CME sites randomly.

Figure 2.2: Computational analysis reveals actin network assembly at the late stage of CME

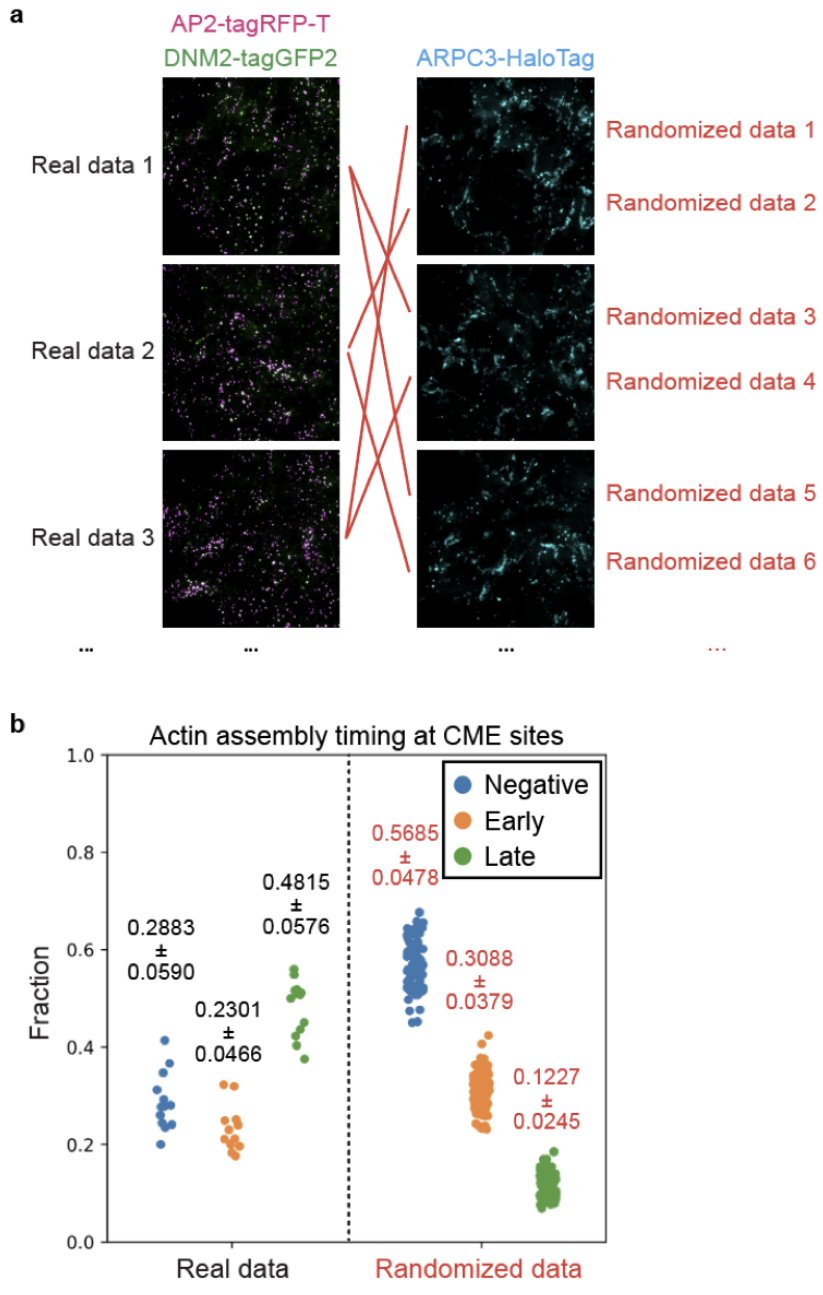


Figure 2.2 legend: a. We generated a randomized data set by pairing ARPC3 images with AP2 and DNM2 images from an unrelated movie. b The fraction of CME sites without actin assembly (Negative), with early actin assembly (ARPC3 signal disappears before DNM2 signal peak), and with late actin assembly (ARPC3 signal overlaps with DNM2 signal peak) were calculated for real data movies and randomized control movies (a). In the randomized dataset, we detected early "assembly" of actin in a similar fraction of CME events as in the real data set, indicating that presence of actin detected early in CME is very likely an artificial association caused when nearby actin structures overlap with CME sites by chance. Mean and standard deviation are shown on the graph.

Figure 2.3: Computational analysis of ARPC3 positive CME sites reveals asymmetric actin network assembly at the late stage of CME

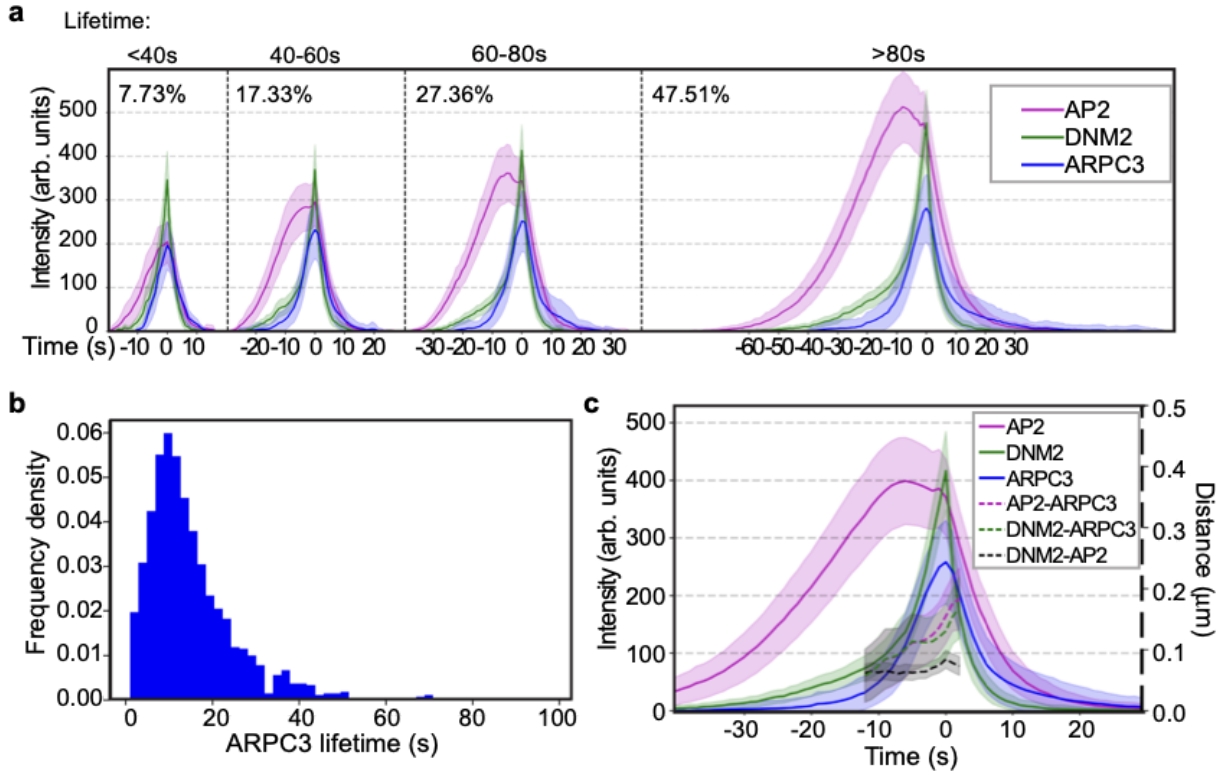


Figure 2.3 legend: a Averaged intensity vs time plots of cohorts of CME sites with late ARPC3 assembly in ADA cells. Events are grouped into cohorts by the lifetimes of AP2 and aligned to the frames showing the maximum DNM2 intensity (time = 0 s). Percentage of the number of the CME sites in each cohort is shown next to the plot. b Histogram of ARPC3-mediated actin network assembly duration. The assembly duration is measured from the first frame of the ARPC3 signal to the presumed scission time (the peak of DNM2 signal). Source data are provided in the Source Data file. c Averaged intensity (solid lines) and distance (dashed lines) vs time plots of ARPC3 positive CME

sites in ADA cells. Events are aligned to the frames showing the maximum DNM2 intensity (time = 0 s). Distance between centers of two signals are shown from -10 s to 3 s when DNM2 and ARPC3 signals are relatively high. a–c N = 1,385. a, c Error bar: ¼ standard deviation.

Our live-cell analysis allowed the timing of branched actin network assembly to be compared to the scission timing, and the spatial offset between the clathrin coat and the associated actin network to be determined. Super-resolution imaging of yeast CME sites suggested that actin and actin nucleators localize symmetrically in a ring around CME sites, and computational modeling suggested that an asymmetric actin arrangement would not provide sufficient force for the membrane invagination during yeast CME[33]. In contrast, in mammalian cells, which require less actin force production during CME, imaging of fixed cells suggested that actin structures associate adjacent to apparent flat clathrin coats[34,35]. However, these studies proposed that at the later CME stages the actin structures become larger and more symmetric to provide sufficient force for membrane deformation and scission. Surprisingly, in our live cell studies designed to highlight sites of new actin assembly, we observed off-centered branched actin networks at CME sites throughout even the latest CME stages. Furthermore, most ARPC3-positive CME sites accomplish scission within 30 s of the initiation of ARPC3 recruitment (Figure 2.3 b). The actin networks (ARPC3) we observed were off center from the coat (AP2) and neck (DNM2) signals by approximately 150 nm at the time of vesicle scission (Figure 2.3 c).

Imaging fluorescent beads using the same settings indicated that the displacement is not an artifact caused by misalignment between different imaging channels (Figure 2.4 a, b). By further analyzing fluorescent bead images, we concluded that chromatic aberration contributes only Given the temporal separation between channel acquisition and the movement of AP2 spots, it was important to assess whether the spatial separation between channels can be attributed in part to an imaging artifact caused by puncta movement between subsequent channel acquisitions. When we measured the average movement of AP2 spots leading up to scission, we found that over 95% of the events had AP2-ARPC3 separations that exceed the frame-to-frame motility of AP2 (Figure 2.4 c). Also, the puncta positional uncertainties indicated by the standard deviations determined when measuring the fitted position of AP2, range up to 40 nm, which is less than the determined displacements. Therefore, we utilized the AP2-DNM2 separation, which is expected to be small, as the basis for comparison to the AP2-ARPC3 and DNM2-ARPC3 separations (Figure 2.3 c). These results further support our conclusion that branched actin networks assemble asymmetrically at CME sites through the time of scission. This observation is consistent with the observation that ring-shaped actin structures at clathrin coats were rarely observed in the

high-resolution, live-cell imaging reported in a previous study[6]. In total, these live-cell data suggest that in mammalian cells, asymmetric actin network assembly can provide enough force to assist membrane deformation and scission during the late stages of CME.

Figure 2.4: AP2-ARPC3 separation is not due to imaging artifacts

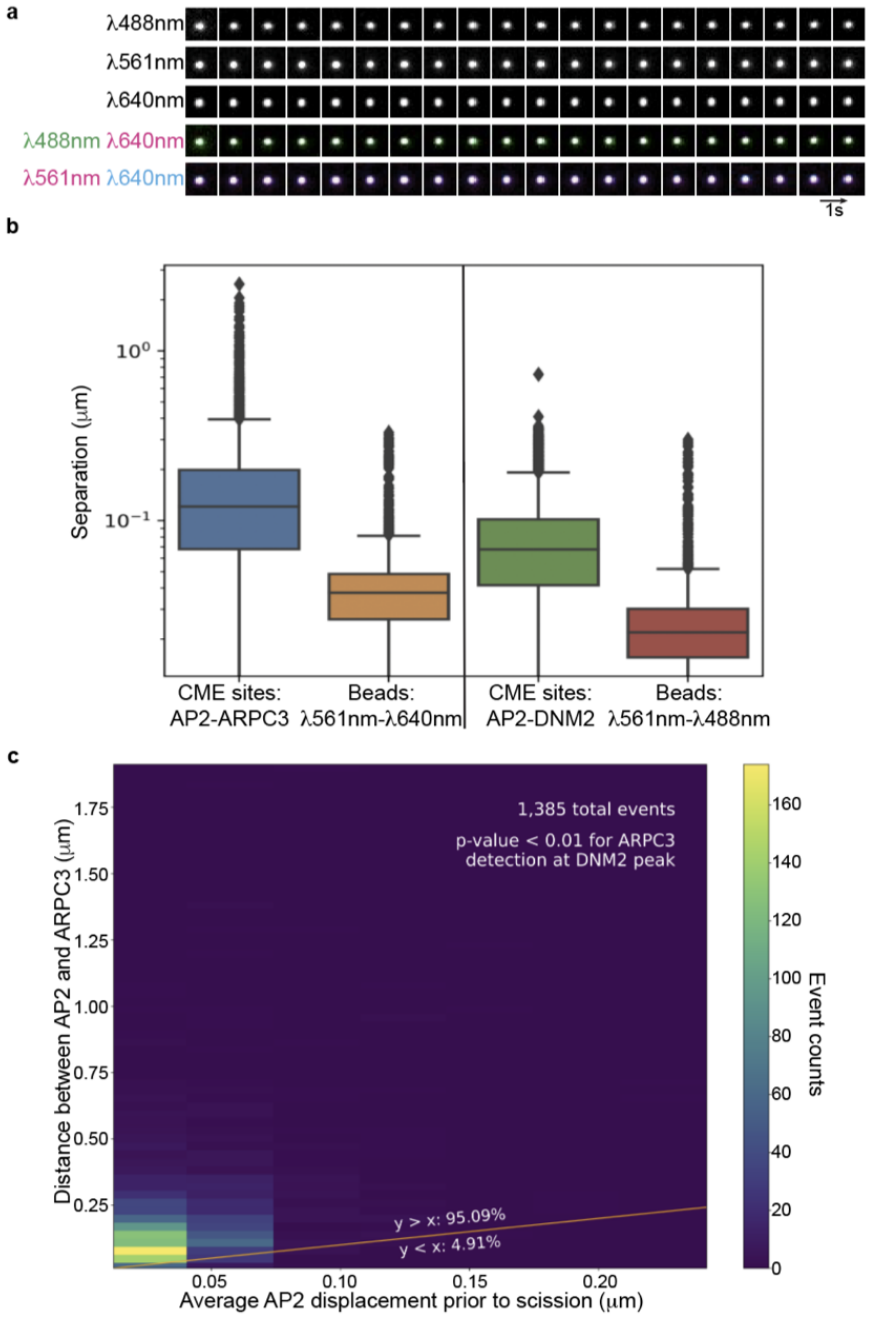


Figure 2.4 legend: a Montage from a TIRF movie of a multi-fluorescence bead. Size of field of view: 2μm x 2μm. Intervals: 1sec. b A boxplot of inter-channel distances (note:

y-axis is base-10 log scale) between centroids of either CME sites or beads to highlight that the observed separation between marked proteins exceeds the measured chromatic aberration. Beads were compared between the same pairs of channels as used to image the tagged proteins. Box plot elements: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. c A heat map graph of distance between AP2 and ARPC3 signals before scission, and average AP2 frame to frame displacement within 6 seconds before scission. Over 95% of the CME events present larger AP2- ARPC3 separation than AP2 displacement. N= 1,385.

Asymmetric branched actin networks facilitate CME at stalled sites

To gain additional insights into the function of this asymmetric actin network assembly, we quantitatively compared kinetics of CME events with or without ARPC3 recruitment. We observed that about 30% of CME events are completed in the absence of detectable actin assembly (Figure 2.2 b), which is consistent with the hypothesis that in mammalian cells actin assembly is required for CME only under certain conditions, such as relatively high membrane tension or specific cargo internalization, which can vary regionally within cells[26,27,29,30,36]. Consistent with the possibility that unfavorable conditions such as increased membrane tension, might stall membrane deformation during CME[24,26,28–30,37–39], CME lifetimes were markedly longer for ARPC3 positive events compared to the ARPC3 negative events (Figure 2.5 a). In addition, when the AP2M1 intensity vs time profiles were compared between ARPC3 positive and negative CME sites, a plateau, which lasts for approximately 10 s, was observed for the ARPC3 positive events (Figure 2.5 b). Based on these observations and previous experimental and computational modeling data[24,26,27], we propose that this plateau in branched actin-positive CME events represents stalled membrane bending due to an unfavorable local membrane environment.

We next tested the hypothesis that the asymmetric actin network might affect the lateral movements of endocytic coats on the plasma membrane. Interestingly, the ARPC3 positive CME sites showed significantly slower, but greater directional lateral movement before vesicle scission compared to the ARPC3 negative CME sites (Figure 2.5 c, d). After scission both ARPC3 positive and negative vesicles showed fast, apparently random movements (Figure 2.5 c, d). These data suggest that the asymmetric actin can stabilize the forming endocytic coat while pushing it in the plane of the plasma membrane with a lateral directional force.

Figure 2.5: Actin positive CME sites show distinct dynamics

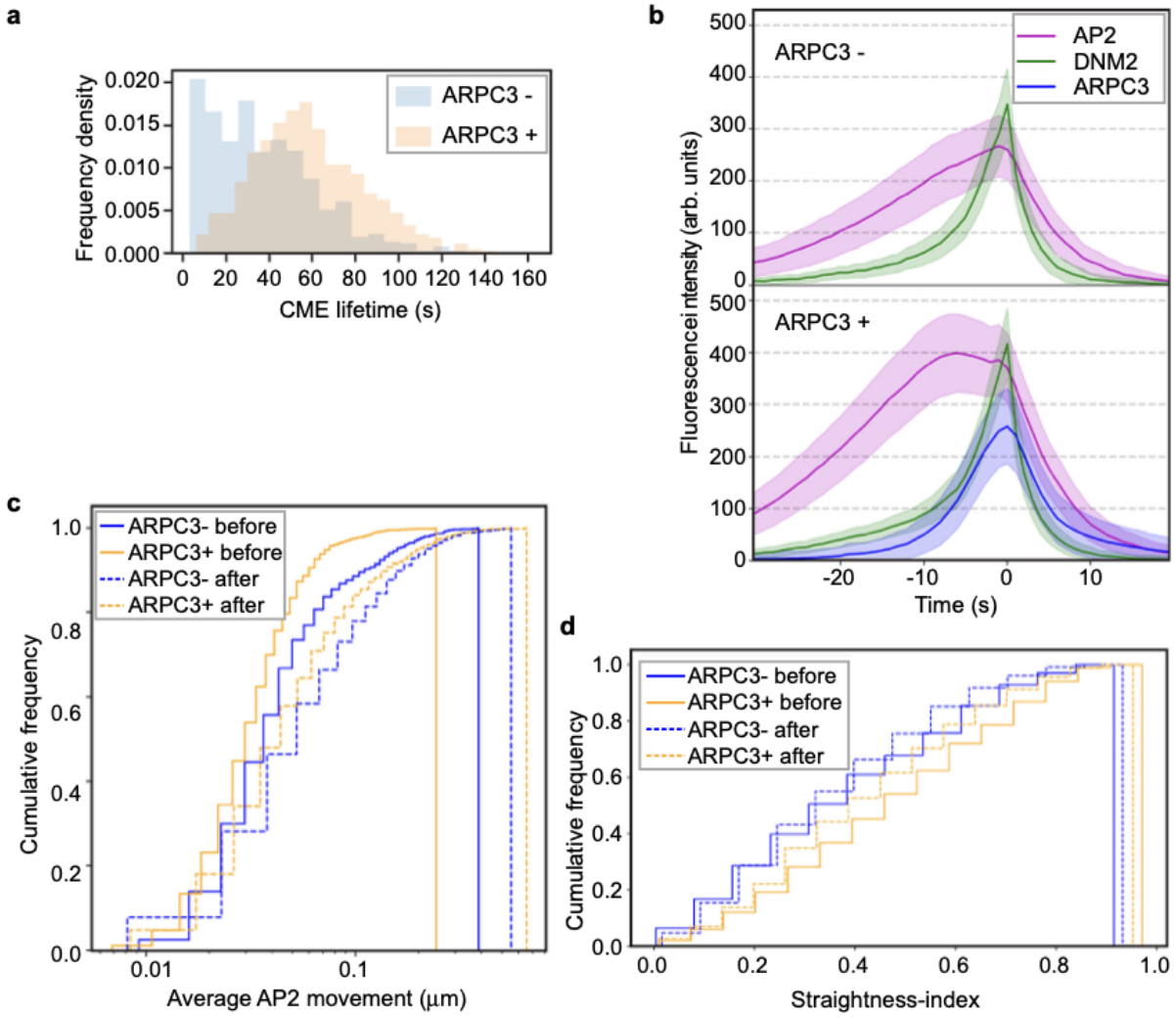


Figure 2.5 legend: a Histograms of ARPC3 negative (blue) and positive (orange) CME lifetimes. CME lifetime is measured from the first frame of the AP2 signal to the presumed scission time (the peak of DNM2 signal). ARPC3 positive CME events have longer lifetimes. p-value from two-sided Welch's t-test: 2.11e-76. b Averaged intensity vs time plots of ARPC3 negative (top) and positive (bottom) CME sites in ADA cells. Events were aligned to the frames showing the maximum DNM2 intensity. Error bar: 1⁄4 standard deviation. c Lateral motility of ARPC3 negative (blue) and positive (yellow) CME sites before (solid line) and after (dashed line) vesicle scission. ARPC3 positive CME sites move slower than ARPC3 negative ones. p-value from two-sided Kolmogorov-Smirnov test: ARPC3 + before vs ARPC3- before: 1.27e-14, ARPC3 + after vs ARPC3- after: 5.18e-08. d Straightness-index of ARPC3 negative (blue) and positive (yellow) CME sites before (solid line) and after (dashed line) scission. The straightness-index is defined by the ratio between the sum of frame-to-frame distances to the end-to-end distance of a single event's trajectory, where a perfectly straight- lined trajectory would have an index of 1. APRC3 positive CME sites move with a more

straight trajectory. p-value from two-sided Kolmogorov–Smirnov test: ARPC3 + before vs ARPC3- before: 1.07e-11, ARPC3 + after vs ARPC3- after: 2.06e-6. a–d ARPC3 -: N = 840, ARPC3 + : N = 1,385.

To test the function of actin network assembly on CME, we treated the cells with an Arp2/3 inhibitor, CK666. We observed CME dynamics immediately after the treatment to minimize non-specific side effects that might be caused by prolonged actin assembly disruption. Even with moderate inhibition of actin network assembly at CME sites, indicated by reduced ARPC3 intensity (Figure 2.7), we detected a small but significant increase in CME lifetimes (Figure 2.6). Additionally, we tested whether hypotonic media that elevates plasma membrane tension via cell swelling would affect ARPC3 dynamics. We found an elevated ARPC3 response at actin positive CME sites, as well as a smaller coat indicated by a global decrease in AP2 signal (Figure 2.8). These results are consistent with results from previous studies in SKMEL cells[27,32] and supports the hypothesis that Arp2/3-mediated actin assembly facilitates CME.

Figure 2.6: CK666 treatment results in longer CCP lifetimes

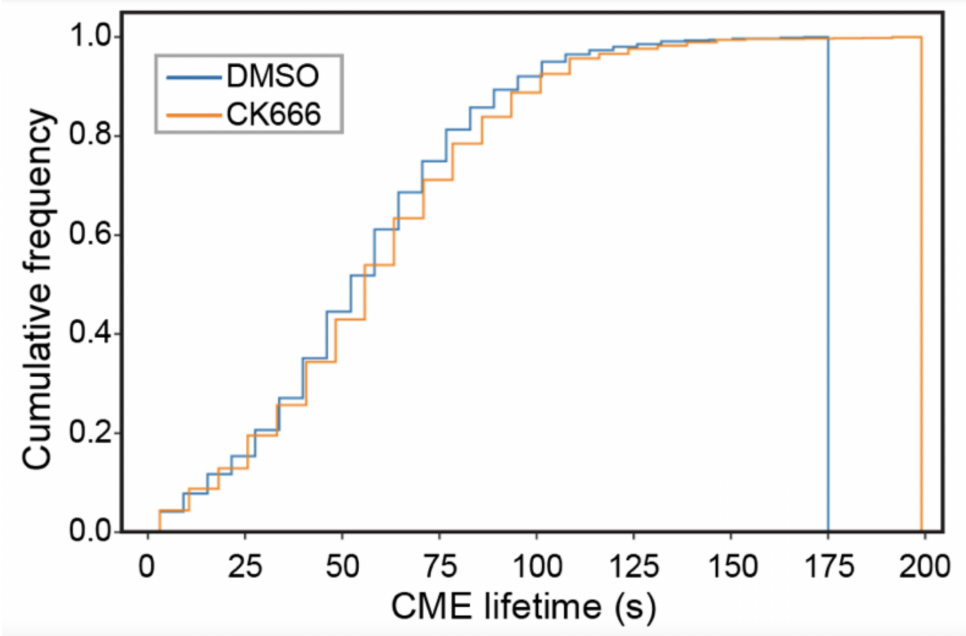


Figure 2.6 legend: A ECDF plot of CCP lifetimes (AP2 initiation to dynamin peak/scission) of control and CK666 treated cells.

Figure 2.7: ARP2/3 response at CCPs after ARP2/3 inhibition via CK666

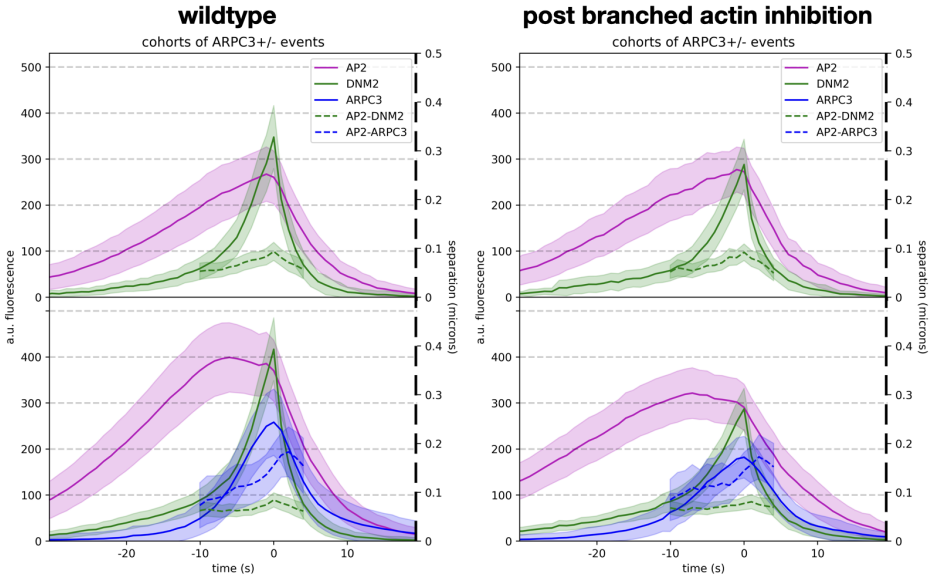


Figure 2.7 legend: Left panel: control (DMSO) cells, top: ARPC3- CCPs, bottom: ARPC3+ CCPs displaying averaged intensities and channel-to-channel separations. Right panel: CK666 treated cells showing decreased ARPC3 signal at CCPs.

Figure 2.8: ARP2/3 response at CCPs after cell swelling via hypotonic shock

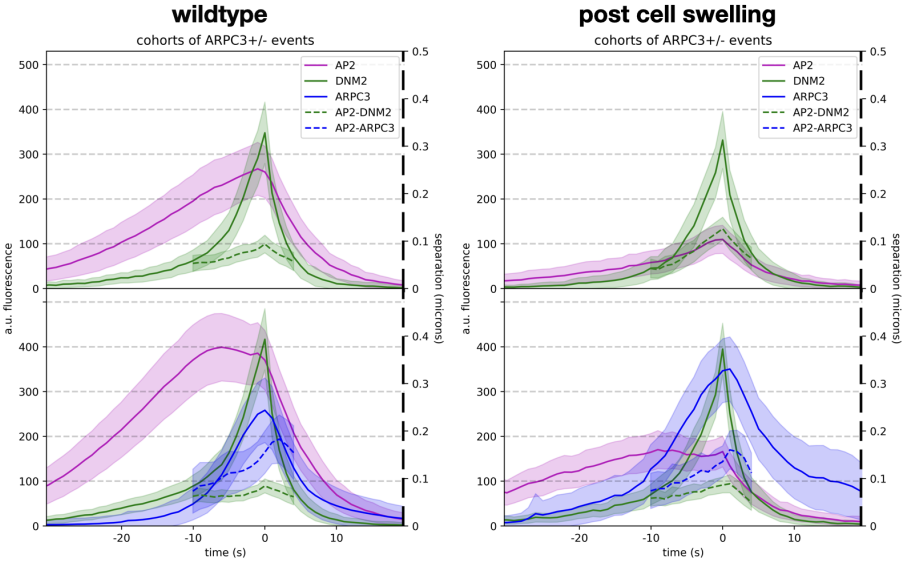


Figure 2.8 legend: Left panel: control media cells, top: ARPC3- CCPs, bottom: ARPC3+ CCPs displaying averaged intensities and channel-to-channel separations. Right panel: hypotonic media treated cells showing elevated ARPC3 signal at CCPs.

Discussion

Using unbiased analysis of thousands of CME sites in unperturbed live cells, our study demonstrates that in mammalian cells clathrin coat assembly dynamics predict which sites will assemble actin, and show that at apparently stalled sites, actin assembles asymmetrically to facilitate successful vesicle formation.

Based on the data presented here, we propose an updated model for actin assembly at mammalian CME sites in which, beyond global tension-dependent changes in the requirement for actin assembly, highly localized differences give rise to heterogeneity even within the same patch of plasma membrane in the same cell (Figure 2.9): (1) Where the local conditions are favorable for membrane deformation by coat proteins (Figure 2.9), the membrane can undergo flat-U-Ω shape transitions in a relatively short time without actin assembly. When the coat grows large enough to form a Ω-shaped bud, sufficient dynamin can be recruited to perform scission, and there is little delay between coat expansion and scission; (2) Where the local conditions are not favorable, presumably under high membrane tension or other impediments, the coat protein-membrane interaction does not generate sufficient force to curve the membrane (Figure 2.9, lower scenario). Here, extra force generation from actin assembly is required[24,26,30]. Asymmetric N-WASP recruitment activates actin nucleation mostly at one side of the clathrin coat, generating an asymmetric force that pulls the membrane into the cell with a similar action to a bottle cap opener. We speculate that this asymmetrical force contributes to asymmetric membrane deformation at endocytic sites observed by high-speed atomic force microscopy[40] and may act with dynamin[41]3 to twist the clathrin pit to promote scission at the neck. CME events with associated actin assembly have longer lifetimes, likely due to a delay between the end of coat expansion and progress toward vesicle scission, requiring adaptive recruitment of actin regulators followed by actin network assembly and membrane remodeling. This result establishes that site-to-site heterogeneity in actin dependence and involvement can be observed without manipulating actin assembly.

Figure 2.9: An updated schematic model of actin-negative and actin-positive clathrin-coated pits in human cells.

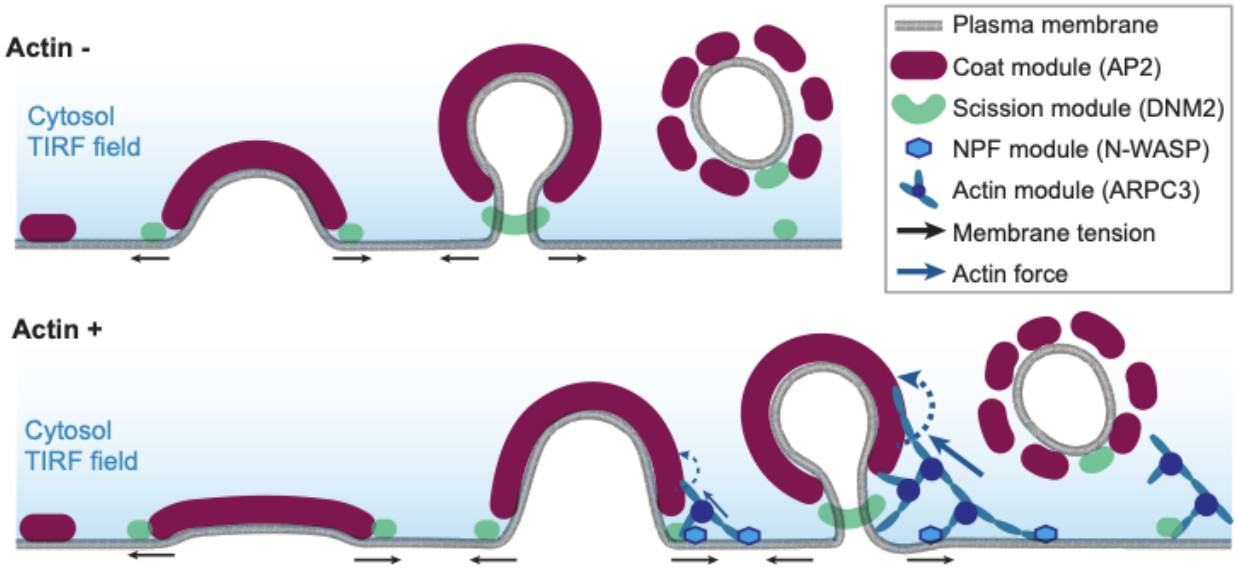


Figure 2.9 legend: Actin assembly is induced at stalled CME sites, where asymmetric forces pull, bend and possibly twist the plasma membrane against membrane tension to drive membrane invagination and vesicle scission.

Our model provides further insights into the basis for inconsistent effects of actin drugs on CME[12,22,26,34,42–46]. Actin plays crucial roles in membrane shaping, cell adhesion, and membrane tension. Global disruption of actin dynamics is expected to dramatically change membrane tension and the available pool of actin and associated proteins and therefore to have both direct and indirect effects on CME. Here, we focused on in-depth analysis of the unperturbed process and detected preference for actin assembly at stalled CME events.

We used genome-edited ARPC3-HaloTag protein expressed at endogenous level as a marker of branched actin networks. This fluorescent protein allowed us to study the spatiotemporal dynamics of Arp2/3-nucleated actin structures at CME sites specifically. A large population of unbranched cortical actin filaments at CME sites has been observed by cryo-EM[47] and distinct effects of cortical actin-based structures on CME have been reported[39]. Previous studies on actin's spatiotemporal dynamics and function at CME sites applied actin probes such as Lifeact, which label actin filaments nucleated by different mechanisms, and which can perturb dynamic actin properties, possibly complicating conclusions reached in those studies[6,34,48,49].

While this study provides evidence that actin assembles asymmetrically at mammalian CME sites, some questions remain open. For example, what is the molecular mechanism that couples recruitment of N-WASP and other actin assembly factors to the stalled CME sites? Many factors, including membrane tension[26,29,30], cell cycle[29], cell

adhesion[28,50] and cargo size and composition[36] appear to contribute to the actin requirement for CME. How do these different factors regulate the actin assembly machinery? Answering these questions in the future will deepen our understanding of dynamic actin assembly regulation in membrane trafficking.

Methods

Cell culture

The WTC10 hiPSC line was obtained from the Bruce Conklin Lab at UCSF. hiPSCs were cultured on Matrigel (hESC-Qualified Matrix, Corning) in StemFlex medium (Thermo Fisher) with Penicillin/ Streptomycin in 37 °C, 5% CO2. Cultures were passaged with Gentle Cell Dissociation reagent (StemCell Technologies, Cat#: 100–0485) twice every week.

Genome-editing

The AP2M1 gene was edited in WTC10 hiPSCs as previously described using TALENs targeting exon 7 of the AP2M1 gene[18]. Both alleles of AP2M1 were tagged with tagRFP-T. The Cas9-crRNAtracrRNA complex electroporation method was used sequentially to edit DNM2 and ARPC3 gene in AP2M1-tagRFP-T genome-edited hiPSCs, as previously described12,18. The same method was used to edit the WASL gene in AP2M1-tagRFP-T/DNM2-tagGFP2 genome edited hiPSCs. S. pyogenes NLS-Cas9 was purified in the University of California Berkeley QB3 MacroLab. TracrRNA and crRNA that target CCTGCTCGACTAGGCCTCGA (DNM2), CCTGGACAGTGAAGGGAGCC (ARPC3) and AGCTCATGGTTTCGCCGGCG (WASL), were purchased from IDT. Gibson assembly (New England Biolabs) was used to construct donor plasmids containing DNM2 5′ homology-ggtaccagtggcggaagc-tagGFP2-DNM2 3′ homology, ARPC3 5′ homology-ggatccggtaccagcgatccaccggtcgccacc-HaloTag-ARPC3 3′ homology, and WASL 5′ homology-HaloTag-agcgatccaccggtcgccaccggatcc-WASL 3′ homology sequences, respectively. Three days after electroporation (Lonza, Cat#: VPH-5012) of the Cas9-crRNA-tracrRNA complex and donor plasmid, the tagGFP2 or HaloTag positive cells were single cell sorted using a BD Bioscience Influx sorter (BD Bioscience) into Matrigel-coated 96-well plates. Clones were confirmed by PCR and Sanger sequencing of the genomic DNA locus around the insertion site. Both alleles of DNM2 and ARPC3 were tagged with tagGFP2 and HaloTag, respectively, and one allele of WASL was tagged with HaloTag in the hiPSC lines used in this study.

Western blotting

Cells were dissociated from the well using Gentle Cell Dissociation reagent (StemCell Technologies, Cat#: 100-0485). Total proteins were extracted by adding 1 ml of cold 10% TCA to the cell pellets, incubated on ice for 30 min, and spun down by centrifuging

at 4 °C, 13400 × g for 10 min. Protein pellets were dissolved in loading buffer (50 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM MgCl2, 5% BME, 5 mM DTT and protease inhibitor) and loaded onto an acrylamide gel for SDS-PAGE and transferred to nitrocellulose membranes for immunoblotting. Blots were incubated overnight at 4 °C with primary antibodies targeting Tag(CGY)FP (1:2000 dilution in 2% milk, Evrogen, Cat#: AB121), HaloTag (1:2000 dilution in 2% milk or 1:1000 dilution in 0.5% milk, Promega, Cat#: G9211), GAPDH (1:100,000 dilution in 0.5% milk, Proteintech, Cat#: 10494-1-AP), respectively, and subsequently incubated in the dark at room temperature for 1 hr with secondary antibodies.

TIRF live-cell imaging

Two days before imaging, hiPSCs were seeded onto Matrigel-coated 4-well chambered cover glasses (Cellvis). Halotag was labeled by JF635- HaloTag ligand[51]. Cells were incubated in StemFlex medium with 100 mM JF635-HaloTag for 45 min and the unbound ligands were washed away by three washes with 5 min incubation in prewarmed StemFlex medium. Cells were imaged on a Nikon Ti-2 inverted microscope fitted with TIRF optics and a sCMOS camera (Hamamatsu). Cells were maintained at 37 °C with a stage top incubator (OKO Lab) in StemFlex medium with 10 mM HEPES. Images were acquired with Nikon Elements. Channels were acquired sequentially at a 1 s interval and 300 ms exposure time over 4 min.

For the CK666 treatment experiments, a 50 mM CK666 (SML0006, Sigma Aldrich) stock solution was prepared in DMSO and kept at −80 °C. 200 μM CK666 solution and 4% DMSO (v/v) solution for a control were prepared fresh in StemFlex medium with 10 mM HEPES and incubated at 37 °C prior to use. Live-cell imaging was performed 1 min after adding an equal volume of medium containing CK666 or DMSO into the imaging chamber to achieve a final concentration of 100 μM CK666 or 2% DMSO.

TIRF image processing and analysis

Events (i.e., tracked diffraction-limited spots) were extracted from cmeAnalysis[133] and processed in Python Jupyter Notebooks. AP2-tagRFP-T was used as the fiducial marker for clathrin-mediated endocytosis in cmeAnalysis tracking experiments. DNM2 was used as a secondary channel to mark vesicle scission and the termination of vesicle formation. ARPC3-HaloTag was tracked separately from paired AP2/DNM2 movies and linked to CCPs downstream of cmeAnalysis to allow for determining discrete ARPC3 nucleation and disassembly events.

Four generalized processing steps were applied to identify clathrin-coated pits (AP2 tracks) with single DNM2 peaks: track feature abstraction, feature dimensionality reduction, event clustering, and DNM2-peak detection. First, tracks that are defined by

fitted positions and intensities for single events were generated using cmeAnalysis. Then, AP2 and DNM2 tracks were decomposed into dynamic features describing the dynamics of the position and brightness for each event. Each tracked event, which was once an arbitrary array of intensities and positions, was now a discrete vector of fixed length. The mapping of each track to discrete features was done to generalize the dynamics of heterogeneous tracked events into a set of interpretable coordinates.

Following feature abstraction, the output array is a 2-dimensional matrix with N rows (N tracked events) and M columns (M discrete features per track). These features were individually scaled to normal distributions to remove the variability in scale and dampen the effects of outliers. For instance, the 'lifetime' feature (AP2 lifetime) ranged from a few seconds to several minutes on a scale of seconds, whereas the 'DNM2-peak fraction' feature (where the DNM2 peak is located within one AP2 event) ranges from 0 to 1. Following feature re-scaling, these events, which each contain over thirty features, were projected to a lower-dimensional space via principal component analysis. The derived clusters were separated using a Gaussian Mixture Model and events were assigned to clusters based on their highest probability of identity to one cluster.

DNM2-positive events represented a distinct cluster of tracks that had detectable DNM2 throughout the event, were long lived, and were below the threshold of motility expected for transient, non-CME-derived clathrin-coated vesicle "visitors" to the TIRF field. To characterize a single-DNM2 peak, DNM2-positive events were surveyed over a range of values set for the minimum DNM2 peak height, width, and peak-to-peak temporal distance. For each peak-defining parameter combination, all DNM2-positive events were categorized as having zero, one, or two or more peaks. After finding single-peaked events in a fixed peak-parameter combination, the lifetime distribution of lifetimes for single peak events was fit to the expected underlying distribution, a Rayleigh distribution, where the best-fitting parameter combination was chosen to identify single-peaked events. Single DNM2-peaked events were kept as CME sites for the remainder of the analysis.

Chapter 3: the dynamics of clathrin-mediated endocytosis in zebrafish embryos

Background

As my collaboration with Meiyan was winding down and our paper was on a path to publication, I began thinking about next steps. I had spent the three years focused on analyzing data and was eager to learn the methods of experimental biology and generate data of my own. Truthfully, I was poking around job boards to see what was out there and figured it would serve me well to be more than a one-trick data scientist: I

needed to understand the limitations of experimental design and interpretation, so I set out to do experiments. This first section outlines how I made the jump to zebrafish developmental biology and highlights how following some serendipity led to a productive stretch of new and unpublished discoveries. I hope somebody will take the seeds in this chapter and build on this data to create a thesis of their own.

I was generally interested in how endocytosis is repurposed as cells differentiate. Work in the Drubin Lab prior to my joining had shown dramatic reconfiguration of key endocytic proteins after a clonal population of embryonic stem cells were differentiated into neural progenitor cells and fibroblasts[42]. Then, I revisited a paper by Joh, my rotation mentor in the Drubin Lab, that described the dynamics of CME at the granularity of different cell surfaces found in polarized cells in intestinal organoids derived from the aforementioned stem cells[52]. Stem cell differentiation into organoids had become a robust commercial enterprise by that point and the lab had an arsenal of stem cell lines with various combinations of interesting CME markers. Together with the fact that our lab was three floors above the burgeoning Advanced Bioimaging Center (ABC), I wanted to study CME in optically complex tissues that were not previously accessible with fluorescence microscopy. The ABC was developing bespoke instruments capable of imaging CME at high spatiotemporal resolution in cell populations which distort light and compromise resolution[53]. Their fix, combining adaptive optics with lattice light-sheet microscopy (AO-LLSM), provided a means to undo the light distortion and image volumetrically with gentle light dosages and high speed. There was only one problem: I hated tissue culture.

Fortunately, the leaders of the ABC, Gokul Upadhyayula and Eric Betzig, had an affinity for zebrafish[53–55], a model organism that could be cultured with antibiotic-free media or using equipment not sterilized with ethanol. Once during my second year of graduate school when Gokul and Eric took me to lunch, I recall Eric saying to me that “[zebrafish are] God’s gift to microscopists”. After doing some reading, it was apparent that zebrafish were an appropriate substitute for stem-cell derived organoids: they have many kinds of stem cells that differentiate according to their design by evolution, they are optically clear during embryogenesis, they breed in high numbers which enables complicated genetics, and their small size allows for imaging of many cells in the window of typical microscope cameras. Unfortunately, at that time I was neck-deep in CME analysis mode and had no time for zebrafish.

Some time later, I attended a faculty candidate seminar by Ian Swinburne, a collaborator of Gokul and Eric who had pioneered some of the seminal pieces required to make zebrafish advantageous for imaging[56,57]. Together, the three of them had paved the foundation of this chapter of my thesis with an experiment involving CME: Ian

injected a chimera of clathrin light chain A with moxNeonGreen (clta-mNG) as mRNA into zebrafish embryos and the AO-LLSM revealed the dynamics of clathrin in multiple differentiated cell types with striking differences between them[53]. There was only one other problem: the clathrin was overexpressed. My training in the Drubin Lab had taught me the perils of overexpression[7,48] such as mistimed protein recruitment, and I was still excited to find a way to leverage the power of genome-editing to study endogenous copies of clathrin in a living animal. Fortunately, David told me Ian had accepted an offer to join the department and that he was excited by the prospect of me studying CME in a living animal. I reached out to Ian to see if this was possible and we decided to start a collaboration.

Ian told me that he had previously attempted to make a genome-edited clathrin strain, but the efficiency was low and work was required to make it scalable. In cell culture, efficiency is a bit of an uninteresting point considering that you are able to simultaneously edit millions of cells and then sort cells via selection (e.g. drug resistance) and then by automated cell sorting of fluorescent clones. I was almost scared away from pursuing this by Gokul who warned me of the woes of screening embryos one-by-one without the luxury of numbers and automated cell sorting afforded by cell culture specialists. To summarize his concerns: zebrafish required manual microinjection delivery of editing materials into 1-cell embryos, a transient state that quickly cleaves into two cells. Then, subsequent raising of candidate “Founders” is required to test for successful germ-line editing. Founders are difficult to raise because they carry with them unintentional mutations from the editing process[58] and must survive in good enough health to have sex as adults to pass properly edited gametes onto offspring. The offspring “F1s” require one-by-one manual genotyping (surgical tissue removal as adults) to clonally isolate an edit of interest since the Founders’ germ-line could have a heterogenous population of editing events. Fortunately, I was set on figuring out how to get zebrafish genome-editing more than one-off happenings and more of a mainstay to study cell biology in living organisms[59].

In this chapter I describe my work of the past two years to 1) get work out zebrafish editing, 2) generate strains to study endocytosis and other cell dynamics, and 3) study iron internalization in developing red blood cells. Going into this project, I initially set out to generate strains to study the following questions: What tissues use CME constitutively? How does CME change during the course of differentiation? Therefore, I needed strains that unambiguously labeled CME sites.

Genome-edited zebrafish can be used to understand how a pathway is remodeled during development in diverse tissues such as the ones shown in Figure 3.1, where a fish fully labeled for clathrin light chain A with moxNeonGreen (clta-mNG) can be

studied in cells ranging from tissues found in skin, brain, muscle, and blood. The fish presented in this chapter have the potential to reveal novel cell biology that sidesteps the need for selecting good transgenic alleles or over-expressing fusions of interest.

Figure 3.1: example of a genome-edited fish showing moxNeonGreen fully labeled labeled endogenous clathrin light chain A

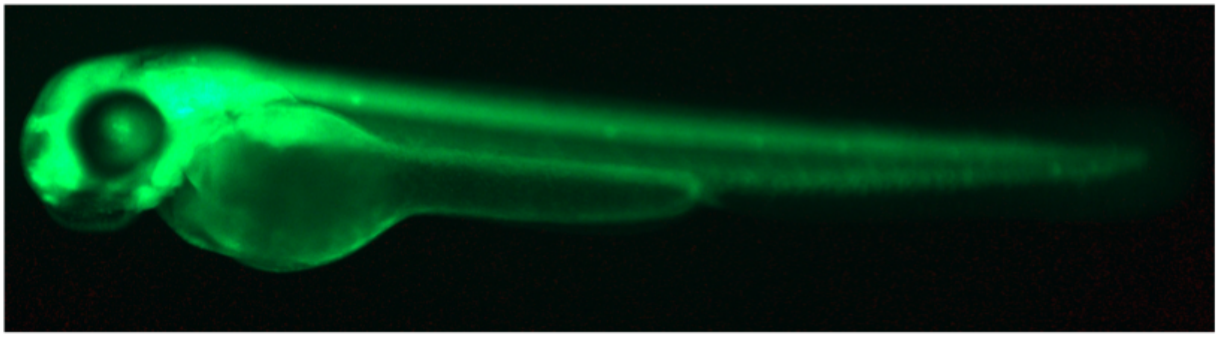


Figure 3.1 legend: a 60 hpf embryo derived from a homozygous clta-mNG (including Casper) in-cross displaying ubiquitous labeling across embryo. Embryo is about 3 millimeters in length.

I opted initially to not create a Dnm2 strain because Dnm2 is duplicated in zebrafish[60]. Ian had already created some of the reagents to label clathrin light chain A (clta) so I sought to label the AP2 complex. Rather than use a scission or other late-stage, low copy-number marker (Auxilin/GAK) to identify CME sites, I proceeded towards a two-color line combining clathrin and AP2 that would be sufficient to identify CME sites[18]. Unfortunately, the mu subunit of the AP2 complex (as used in Chapters 1 and 2) is also duplicated in zebrafish[61]. I nonetheless initially attempted to tag the ap2m1a subunit but could not see fluorescence in the F0 generation (data not shown). I abandoned that fusion and instead resorted to the sigma subunit that other labs used[62] which was fortunately not duplicated in zebrafish (ap2s1). I was also interested in whether there would be stark differences in Arp2/3 recruitment based on developmental time points or cell types, so I aimed to create an ap2s1/arpc3 strain.

As I got better at zebrafish editing, I realized it could be productive to image a cargo. Adhesion complexes are known to be remodeled by CME[63] so I set out to find suitable fusions. At the time, the Allen Institute for Cell Science (AICS)[4] and others[64,65] had created cell culture lines with a beta-catenin (ctnnb1) fusion and demonstrated its functionality. Beta-catenin is a component of adherens junctions, adhesion complexes that mediate cell-cell contacts at the plasma membrane, and also has the remarkable ability to translocate into nuclei to act as a transcription factor[66] that is triggered during canonical WNT signaling[67]. I am thankful to members of the AICS, Brock Roberts and Irina Mueller, who answered my questions about targeting beta-catenin. Ian was

supportive of my mission to create this adhesion-labeling strain for studies of CME since he was interested in WNT signaling. I came across some potentially understudied cellular structures during my imaging of my moxNeonGreen-ctnnb1 (mNG-ctnnb1) line. I unfortunately never got the opportunity to carefully image AP2/clathrin with beta-catenin.

I was also excited by the possibility of studying other endocytic pathways such as caveolae-mediated endocytosis[68] in parallel with CME. Bob and I had shown that caveolae, unlike clathrin-coated structures, do not preferentially localize to artificially-induced cellular curvatures[69], so I was curious if I could see this *in vivo*. Again, Ian was supportive of this because caveolae are implicated in eye development[70] which his lab studies. My results were confirmatory, showing the expected localization of caveolae in cells such as skin and those in the notochord.

The work described in this chapter would not be possible without my labmates in the Swinburne Lab, which I joined in my third year: Samara who was there from the start, Erin and Nate who joined me fall semester of my fourth year, Laurie who made the beta-catenin line during her first-year rotation with Erin and Nate's help, Evan who joined the next spring semester, and Emilia who joined fall of my fifth year. Together, I named our squad the Squids of the Gene-editing Squad. Because there are squids in a squad. Who edit genes. That is the joke.

Results

Generating zebrafish knock-ins

The first few genome-edited strains I created (ap2s1, clta, arpc3) followed protocols similar to those used by pioneers in the field[71]. Guides for generating knock-ins already exist[72] but do not discuss how to deal with organisms like zebrafish with high rates of single nucleotide polymorphisms (SNPs)[73]. Erin Hong, an undergraduate scholar I worked with, provides an in-depth guide for creating our zebrafish knock-ins in her honors thesis, which will be made available online. Here, I supplement her guide by creating visuals highlighting some small innovations that help with scaling up the cloning required to create multiple knock-ins in parallel that deal with SNPs in local populations of zebrafish.

First, in order to create a knock-in, we seek to trigger homology-directed repair, which requires we know the genomic sequences upstream and downstream of our desired insertion site. To facilitate amplification of the wildtype locus surrounding the insertion site (Figure 3.1 A), Evan, an undergraduate scholar I worked with, created a Python

wrapped for that calls primer3[74] to generate amplification oligos for polymerase chain reaction (PCR) according to a set of rules we created[75]. Briefly, the sequence space is searched for oligos that we empirically found to work well in zebrafish: 22-25 nucleotides in length, 2-3 GC-clamps, ~40-60% GC content, and minimal poly-nucleotide stretches. If these requirements could not be met, the algorithm relaxes the search requirements until it converges on a minimum number of suitable upstream and downstream oligos. Usually, we would attempt a minimum of 3 upstream and 3 downstream oligos and use all combinations (M upstream x N downstream = M*N reactions) with touchdown PCR (e.g. 98 2m| 98 10s, 72-1n 15s, 72 3m| 17x| 98 10s, 55 15s, 72 3m| 35x| 72 10m| 12 inf) until a strong amplification pair was identified (Figure 3.1 B). This step is critical because this pair of wildtype locus amplification oligos will be recycled downstream.

These wildtype locus amplification oligos will be used to create the wildtype amplicon (Figure 3.1 C) that requires sequencing to identify the SNPs found in the parental strain that will be crossed to generate knock-ins. Ideally, a handful of fish are individually housed and identified that can be individually genotyped and sequenced. This amplicon (or amplicons with multiple fish) will be used to identify homology arm sequences (Figure 3.1 D). The homology arms can be cloned off of the wildtype amplicon (less background than the entire genomic DNA pool), and then to facilitate isothermal assembly, overhangs can be added directly to the amplification oligos that anneal directly with the homology arm amplicons (Figure 3.1 E). These oligos should have incorporated mutations to nullify the protospacer adjacent motif (PAM), too.

Then, a standard linearized plasmid backbone with ends matching the homology arms' overhangs can be used to assemble the four pieces into a circular plasmid with isothermal assembly[76]. The circular plasmid (Figure 3.1 G) contains landing sites (stars in Figure 3.1 H) that can be used to linearize the homology arms and insert(s) into an amplicon using a standard set of oligos with 5' biotin moiety[77] (donor amplicon oligos), creating an amplicon that is used as a template for homology directed repair. Following repair, individual clones can be rescreened by pairing the original outside homology oligos (Figure 3.1 C) with oligos that anneal specifically to the insert to create junction amplicons. Once fish are homozygous for the insert, the original outside homology oligos can be used together to amplify the whole locus seamlessly.

Figure 3.2: a simplified cloning strategy to scalably use SNP'y genomic DNA for homology directed repair templates

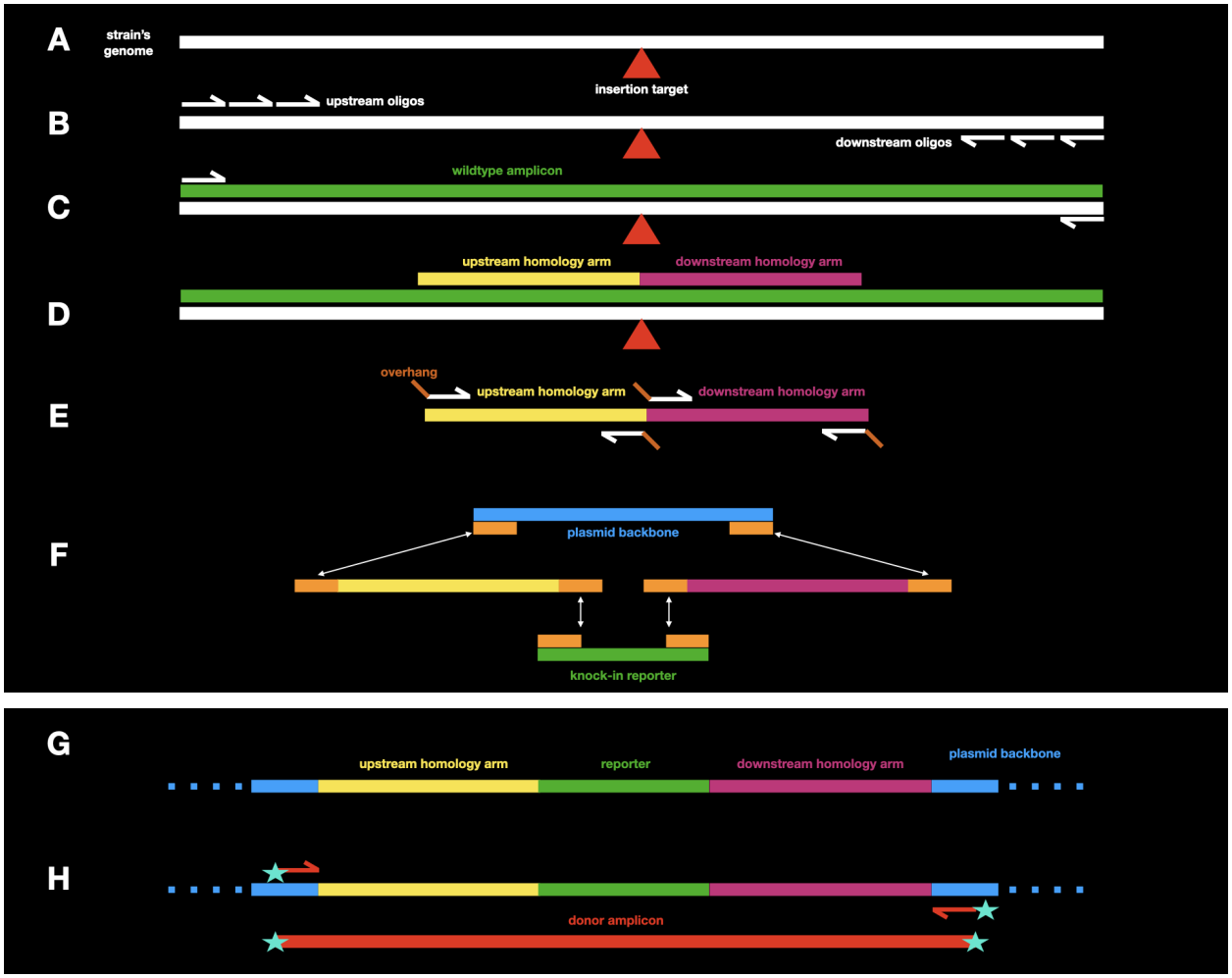


Figure 3.2 legend: A: a reference region found within a sample of whole genome extract that contains a site with a desired insertion point, B: up and downstream oligos are sampled to find a pair that efficiently amplify the region surrounding the insertion site, C: a wildtype amplicon is derived from the region to identify SNPs relative to reference genome and as a template for D: the wildtype amplicon is used as a template to amplify homology arms that contain oligo-added overhangs in E: that correspond to adjacent cloning fragments, F: universal pieces including the backbone, insert, and overhangs from adjacent pieces, G: linear representation of pieces-of-interest from assembled, circular donor plasmid, H: universal backbone contains landing sites for oligos containing 5' biotin moiety (stars) to linearize donor amplicon containing homology arms flanking an insert. Note: objects in the diagram are not to scale.

Note that the linearized plasmid backbone (with ends that are landing sites for donor 5' modified donor amplicon oligos), the overhangs that extend past the homology arms, the knock-in amplicon and its overhang oligos, and the inside-insertion oligos for creating junctions are all potentially universal pieces. I would advise adopting this strategy when creating reagents for more than a few knock-ins in parallel. By creating

homology arms off of the wildtype amplicon from select genomic DNA sources, the issue of nonexistent SNPs should be dealt with. By recycling parts, assembly during cloning can be streamlined.

Finally, during the cloning of junctions (amplicons spanning outside homology arms to inside inserts), sequencing can become cumbersome. Often, I would screen 16 F1s at a time to have 8 males and 8 female candidates to further propagate. To facilitate dealing with many sequencing files, Evan and I came up with a way to rename sequencing files using a template that could be auto generated and organize the sequencing reaction set-up and data interpretation. For instance, we had situations with 16 ap2s1-mNG clones, each with a left and right homology amplicon to sequence, each amplicon requiring 5 oligos to Sanger sequence: e.g. oligos 1-5 for the left junction and oligos 6-10 for the right junction. Evan's code[78] will generate a list of ordered tubes (Figure 3.2 B) that can simply be numbered 1-N (N is 16*2*5 = 160 in example above) when submitting to sequencing and a template will be generated to both order your tubes and then rename files (Figure 3.2 A). Later, clones can be sorted into folders with a reference map as well as corresponding sequencing files. This program dramatically changed how quickly I could assess which clones needed to be culled or differently grouped.

Figure 3.3: organizing sequencing files from repetitious clone screening

Figure 3.3 legend: A: raw files from sequencing returned from tubes labeled 1-N, B: descriptive file names sourced from auto-generated template to organize sequencing set-up and renaming files

The localization and dynamics of organelles in developing zebrafish

In the next sections I highlight some observations of the localization and dynamics of the different markers I engineered in zebrafish. A lot of the data is confirmatory, which is a good thing because it highlights that these fusions preserve protein function, but there are some potentially novel observations made by high resolution fluorescence microscopy via Airyscan microscopy. All of the cell types I describe are classified based on morphology and developmental staging, and they are not confirmed with reporter lines or staining.

Localization of caveolin in developing skin and notochord

Caveolae are membrane invaginations that can traffic vesicles through endocytosis and also sequester signaling molecules[68,79]. Caveolae can be fluorescently labeled by tagging a major scaffolding protein caveolin-1 (cav1). Erin, Nate, and I generated a caveolin1-moxNeonGreen (cav1-mNG) strain and were lucky to find a Founder during a time with substantial difficulty in raising embryos in the zebrafish nursery. The current state of those fish are confirmed genotyped F1s that were the result of the Founder out-crossed to a Casper fish. Caveolae are organized into dynamic puncta as expected in tissues known to have strong expression such as the epidermis and the notochord[80] (Figure 3.4). Interestingly, cav1 localizes to both the apical and basal surface of the epidermis, unlike clathrin and AP2 which are strictly basal (shown later). I think the potential of the line can be harnessed if these fish are crossed with strains known to have defective caveolae effector proteins in order to visualize how the localization and dynamics of cav1-mNG change and contribute to a disease state[81].

Figure 3.4: endogenous caveolin expression in epidermis and notochord

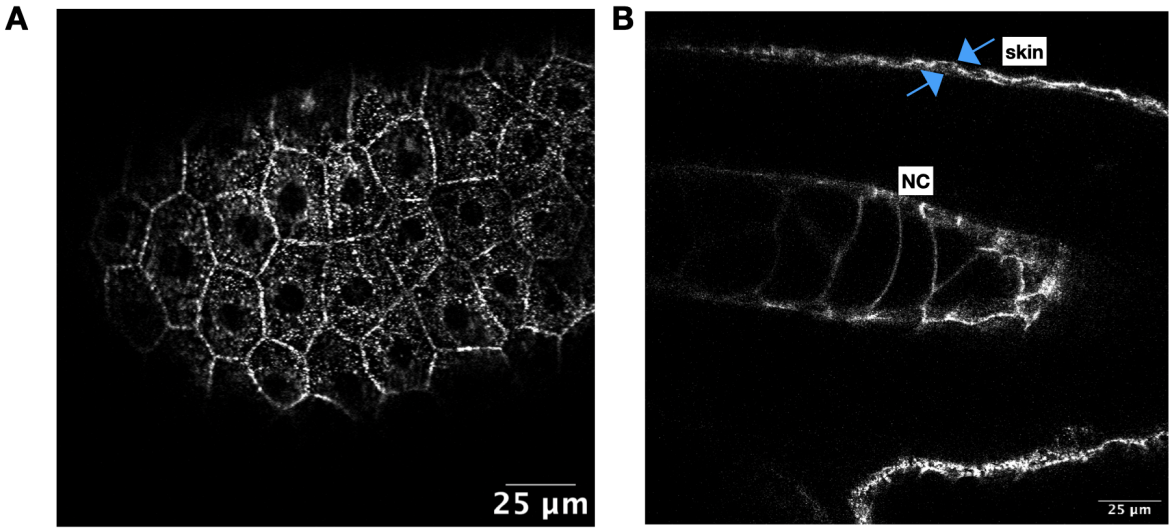


Figure 3.4 legend: A: cav1-mNG (F0 ox Casper) 24 hpf fish shows caveolin enrichment on epidermis, B: caveolin labels both apical and basal surface of epidermis (unlike CME, which is strictly basal), as well as strong enrichment on cells of the swelling notochord.

A beta-catenin line to study adhesion dynamics and canonical WNT signaling

I was fortunate to work with Laurie Wang who started her rotation with me right after I had started testing reagents for generating a beta-catenin line. At that point, I had tested cutting of a guide RNA and Laurie helped assemble a donor plasmid. After much troubleshooting, we ended up switching guides to one that allowed us to edit enough cells while keeping embryos healthy enough to raise. Those babies were miserable to raise and I will spare the details here, but I triumphed and was able to sustain two different Founders with homology-directed repair germline alleles. I tracked one F1, aka Big Chongus, who is the parental clone of all subsequent generations (Big Chongus = Founder #1 ox Casper). I out-crossed her again with Caspers to derive Casper'd mNG-ctnnb1 hets, which I subsequently in-crossed to generate fully homozygous Casper beta-catenin adults.

The data from this line is beautiful, and I think we have just started to scratch the surface of the potential for this line. For starters, we have been able to see the classic view of nuclear beta-catenin accumulation[82,83] during brain development (Figure 3.5).

Figure 3.5: endogenous nuclear beta-catenin accumulation in the developing brain

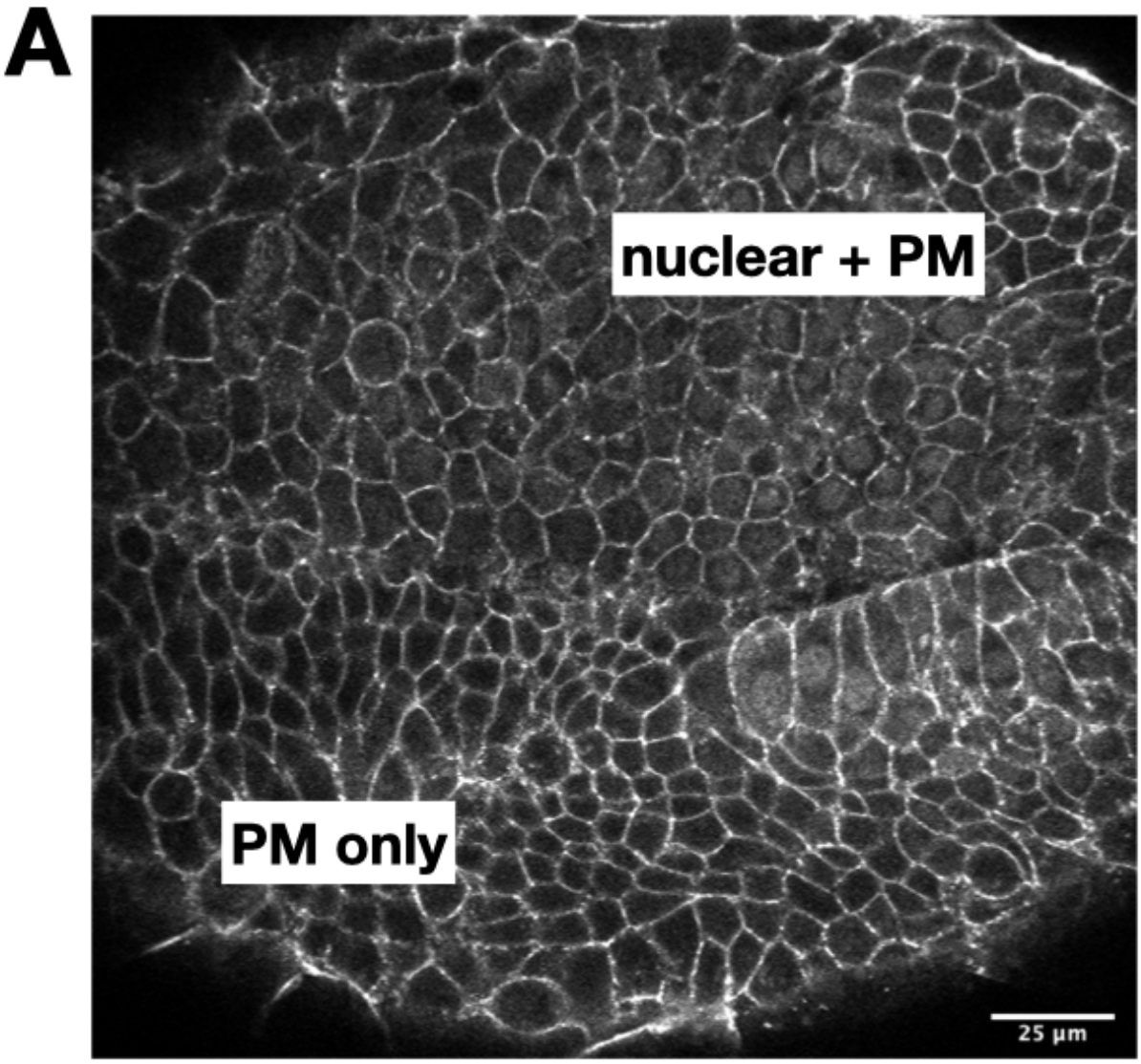


Figure 3.5 legend: beta-catenin (mNG-ctnnb1 from F0 ox Casper) at 13 hpf is enriched at the plasma membrane (PM) via adherens junctions and also accumulates in the nucleus of a subset of cells undergoing canonical WNT signaling

However, the otic vesicle, which is known to be patterned by canonical WNT signaling[84] shows labeling patterns closer to what is observed in cell culture[65] where cells in the dorsal epithelium and surrounding mesenchyme (Figure 3.6 A), as well as cells in the center of neuromasts (Figure 3.6 B) show a filled cytoplasm rather than a strictly-nuclear accumulation as in Figure 3.5. This abundance of intracellular signal is tilted towards cells in the dorsal ear compared to the ventral ear (Figure 3.6 A), confirming evidence that canonical WNT signaling patterns the dorsal ear. There are novel punctate structures at the plasma membrane of this mesenchyme (Figure 3.6 A) as opposed to the uniform labeling found on adherens junctions at the apical surface of skin and neuromasts (3.6 D). Finally, in this same stack, the basal surface of skin shows

a ruffled pattern (Figure 3.6 C) that may be understudied structures formed with beta-catenin that participate in basement membrane adhesion via anchoring fibrils[85].

Figure 3.6: underexplored beta-catenin patterns in the developing ear and surrounding tissues

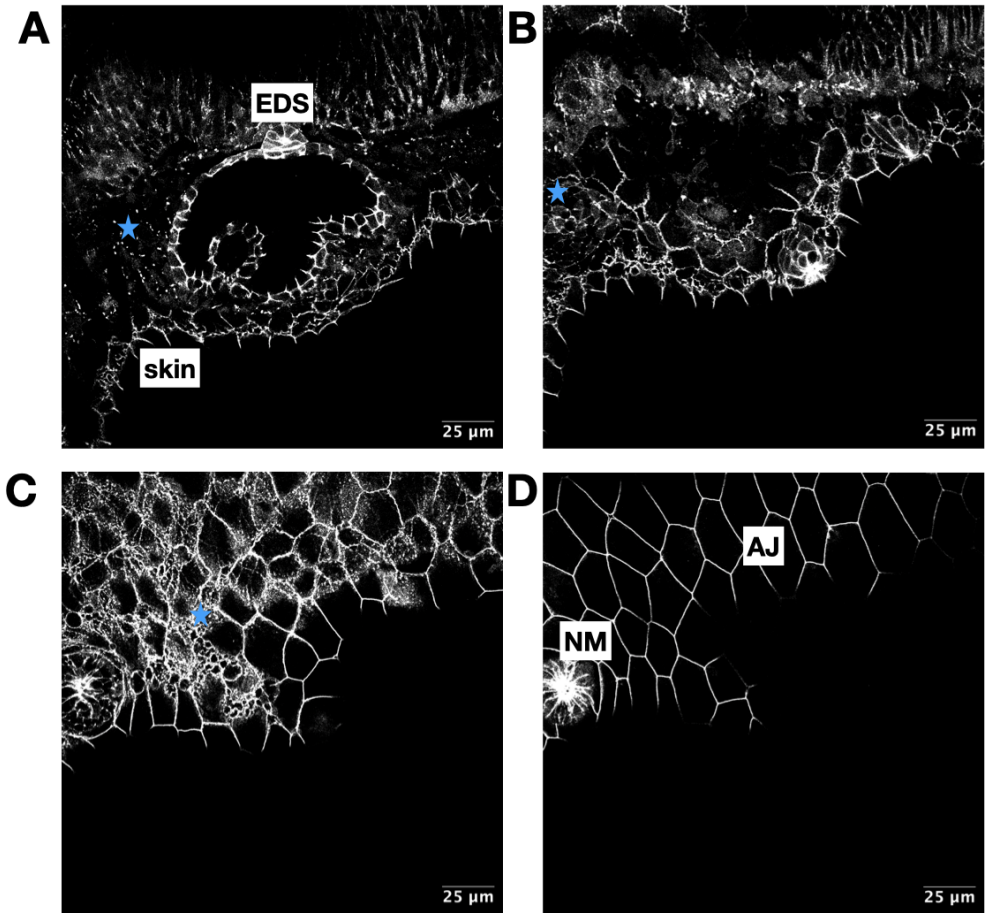


Figure 3.6 legend: beta-catenin-mNeonGreen (mNG-ctnnb1) homozygous (including Casper) in-cross at 50 hpf. Gamma compression was used in these images to facilitate the visibility of structures found in a large dynamic range of intensities in this stack. All images are 2-D slices of the same 3-D Z-stack. A: the endolymphatic duct and sac (EDS) has strong adhesion and intracellular accumulation; higher cytoplasmic signal is found on cells in the dorsal (up) versus the ventral (down) epithelium of the otic vesicle. Mesenchymal cells (blue star) adjacent to otic vesicle have strong intracellular signal but their plasma membranes do not display uniform epithelial labeling but instead organize as punctate structures. B: select cells (blue star) in the center of the neuromast have strong intracellular accumulation of beta-catenin. C: the basal epidermal surface of epidermal cells have enriched signal in the form of anchoring fibrils attached to basement membrane, D: adherens junctions (AJ) at the apical surface of skin and the neuromast (NM)

Finally, there are some potentially novel structures outlined in Figure 3.7. At the basal surface of skin surrounding the yolk (Figure 3.7 A), where there is likely no basement membrane, I have observed finger-like protrusions enriched for beta-catenin that I coined “fingoos”. Occasionally in neurons such as those in the neural tube (Figure 3.7 B) and the brain (Figure 3.7 C), there are large dynamic vesicles that bounce around the cell but appear much smaller than the expected size of the nucleus that I coined “bouncy bois”. I suspect that these compartments are lysosomes containing adhesion complexes that have been internalized during neuron maturation. Additionally, in developing blood (Figure 3.7 D), potentially immune cells because of their membrane ruffles, there is strong intracellular accumulation of beta-catenin, which is known to be important for blood development[86].

Figure 3.7: diverse beta-catenin labeling in skin, neural tube, brain, and blood

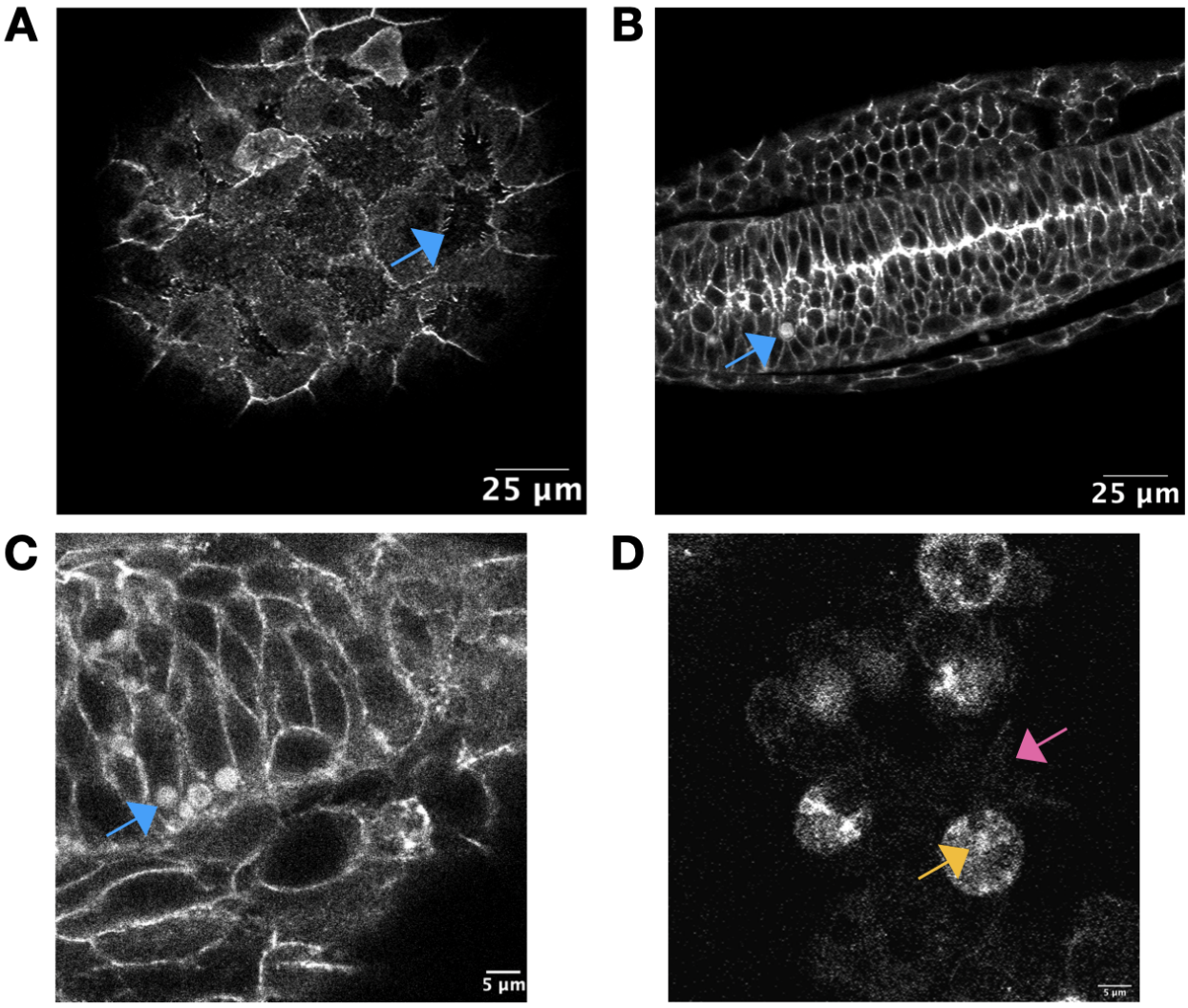


Figure 3.7 legend: A: beta-catenin-mNeonGreen (mNG-ctnnb1) homozygous (including Casper) in-cross. A: 16 hpf, basal surface of epidermis wrapping around the yolk sac shows finger-like projections coined “fingoos” (blue arrow) in an area potentially devoid of basement membrane, B: 16 hpf, occasional large vesicles coined “bouncy bois” (blue arrow) found in neurons along the neural tube without strong nuclear accumulation, C: 24 hpf, dynamic vesicles (blue arrow) in brain neurons that shuttle across cells, D: 24 hpf, strong intracellular compartments (yellow arrow) of beta-catenin in developing blood in common cardinal vein, these cells might become immune cells hinted by large membrane ruffles adjacent to them (magenta arrow)

Arp2/3 localization in a living embryo

By generating a homozygous arpc3-mNG line, I was able to observe, for the first time, endogenous Arp2/3 in a living organism which could provide insight into various force-generating requirements found in cells across the zebrafish. For sample, known Arp2/3 enriched structures such as skin microridges[87] (Figure 3.8 A and Figure 3.10) show labeling on the apical surface, as well as potentially under-explored bands of Arp2/3 that wrap the circumference of the apical skin surface (Figure 3.10) that I coined the “Arp2/3 fuel station”. Dense punctate structures can be found in diverse cell tissues such as the brain (Figure 3.8 A), endoskeletal disk (Figure 3.8 B), cloacae (Figure 3.8 C), the apical surface of hair cells and the endolymphatic sac (Figure 3.9 and B), as well as crawling immune cells (Figure 3.9 B) that are chock full of Arp2/3.

Figure 3.8: Arp2/3 labeling in brain, skin, chondrocytes, and cloacae

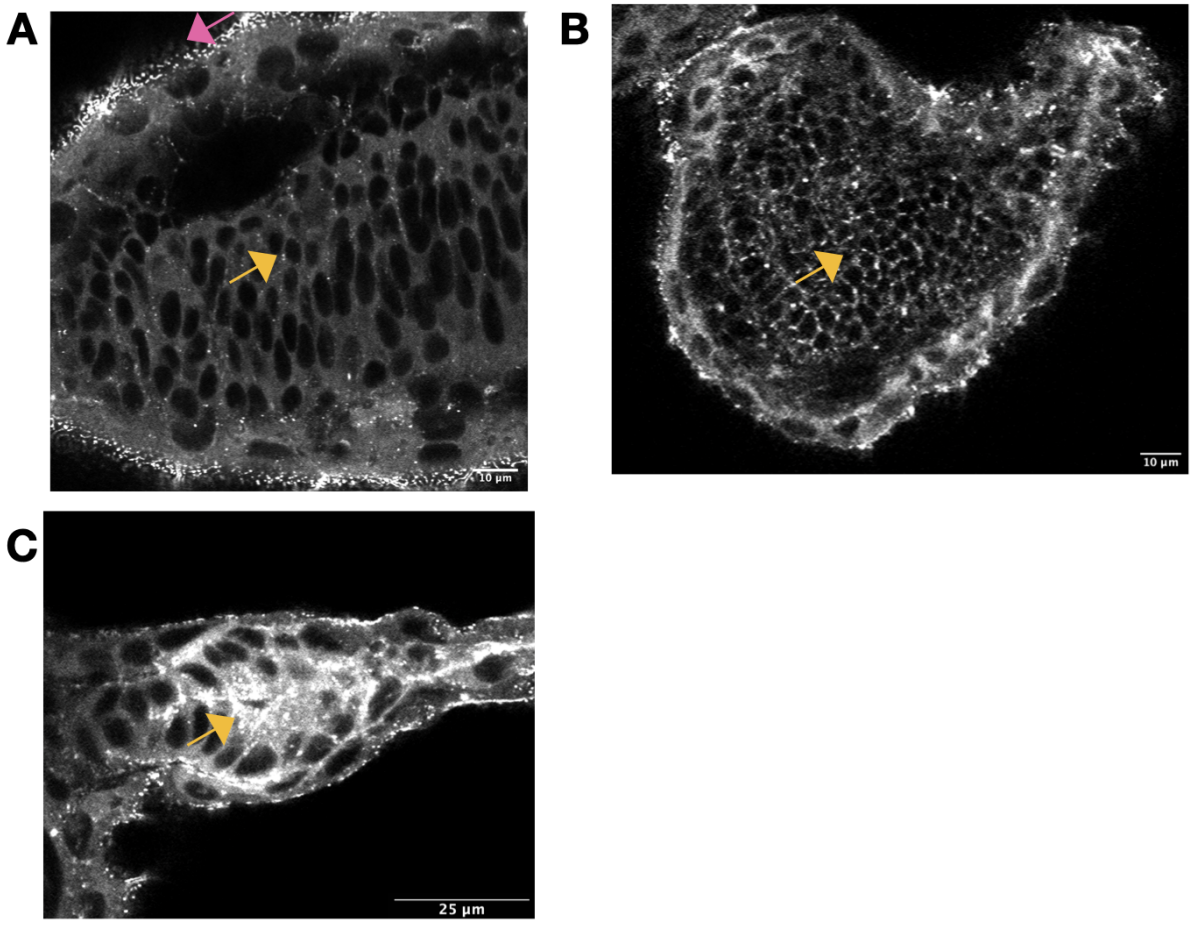


Figure 3.8 legend: Arp2/3 labeled by arpc3-mNG (homozygous in-cross): A: 30 hpf, punctate structures found in brain neurons (yellow arrow) vs pegs and ridges found in apical surface of skin (magenta arrow), B: 50 hpf, bright punctate structures in endoskeletal disk (likely chondrocytes, yellow arrow), C: 30 hpf, strong accumulation of Arp2/3 in the developing cloacae

Figure 3.9: strong Arp2/3 enrichment on the apical surface of hair cells and EDS, and crawling neutrophils

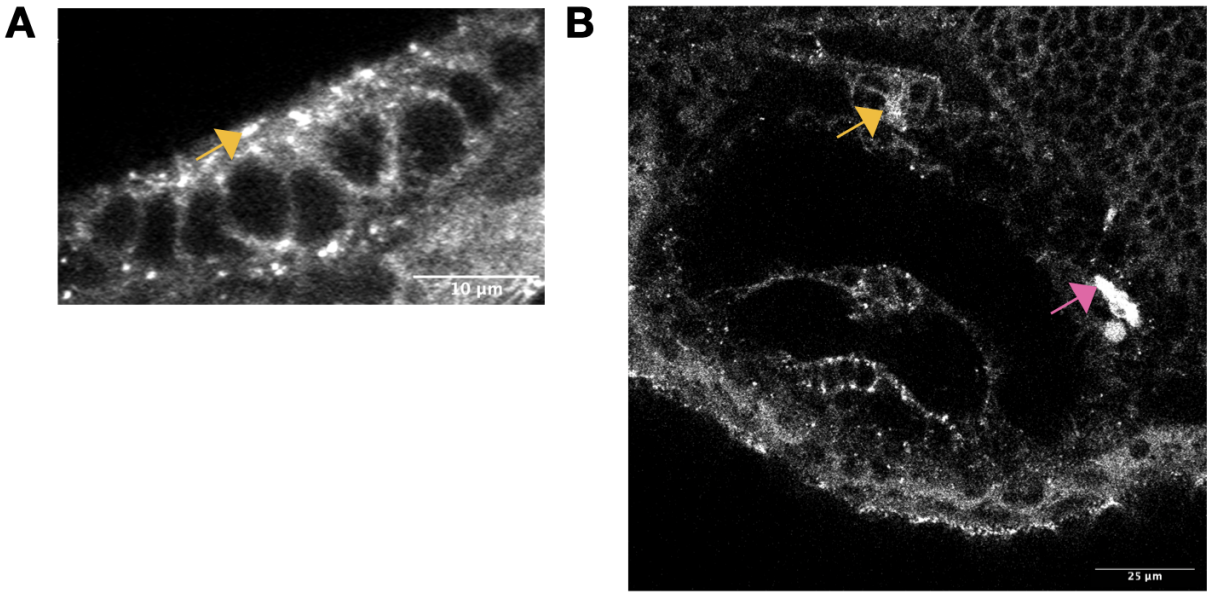


Figure 3.9 legend: Arp2/3 labeled by arpc3-mNG (homozygous in-cross), A: 50 hpf, more dense Arp2/3 at apical surface (yellow arrow) of hair cells compared to basal surface), B: 55 hpf, endolymphatic sace (yellow arrow) has strong polarized accumulation at apical surface compared to basal and lateral surfaces, occasionally immune cells (magenta arrow) crawl around otic vesicle and have strong intracellular pools of Arp2/3

Figure 3.10: Arp2/3 labeling on apical surface of epidermis in microridges

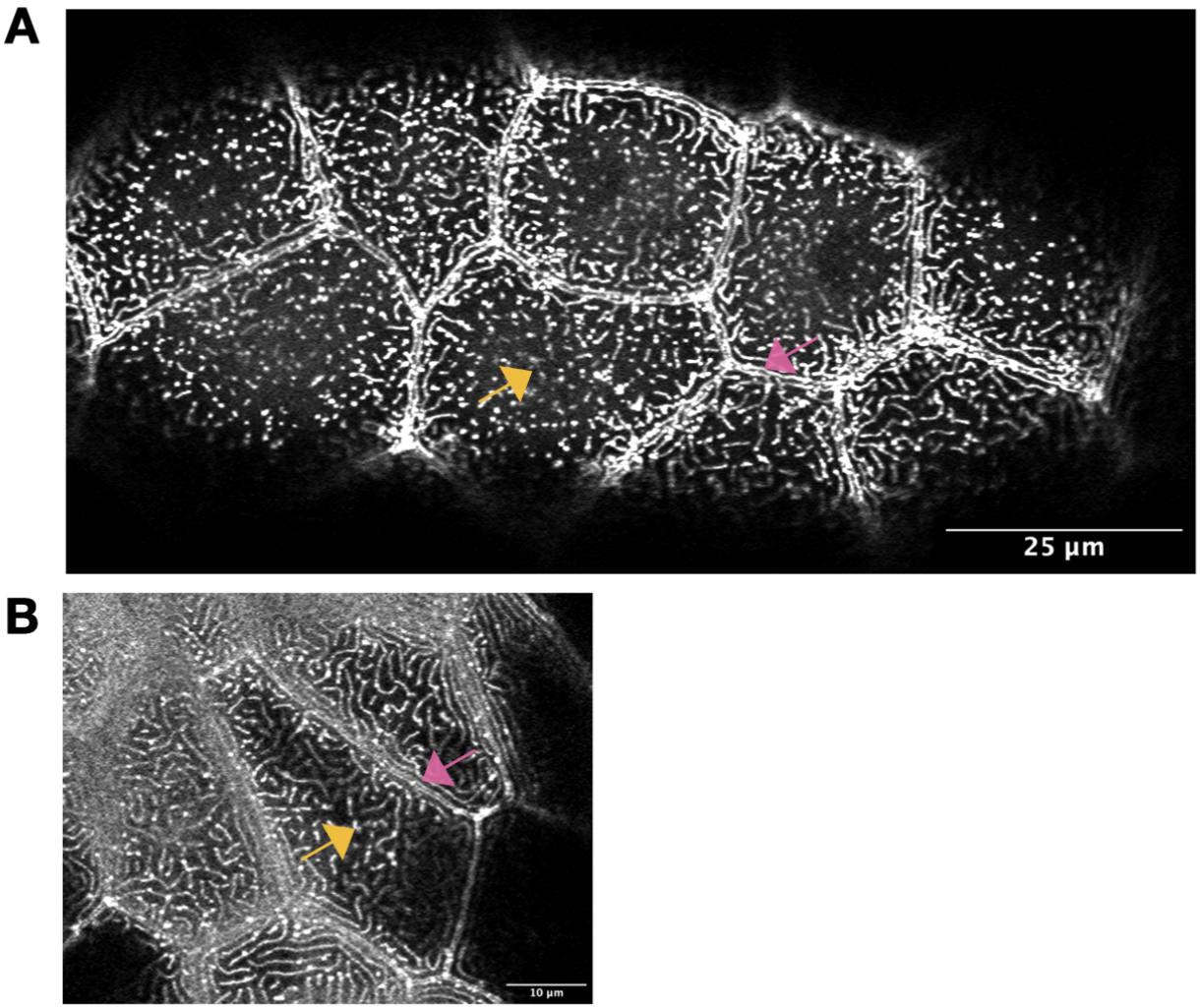


Figure 3.10 legend: Arp2/3 labeled by arpc3-mNG (homozygous in-cross), A: 24 hpf, pegs and microridges (yellow arrow) and a sharp band at periphery of apical surface of skin (magenta arrow) forming the here coined “Arp2/3 fuel station”, B: 50 hpf, pegs and microridges (yellow arrow) and the Arp2/3 fuel station (magenta arrow)

The topic of actin’s importance in non-yeast endocytosis has been debated because of the facilitatory role observed in cultured mammalian cells[26,34,44–46]. By combining my arpc3-mNG line with my ap2s1-mSI line, I was able to find co-localized AP2 structures with Arp2/3 in the pectoral fin (Figure 3.11) which has a strong enrichment for Arp2/3. This may be because of the swelling chondrocytes[88] that have higher membrane tension with higher force-generation requirements for vesicle internalization, and also because of the superficial nature of the tissue which reveals brighter structures that are not aberrated unlike deeper cells. This data presents the first direct evidence of branched actin participating in CME of cells not grown on artificial substrates.

Figure 3.11: Arp2/3 co-localization with AP2 in the pectoral fin

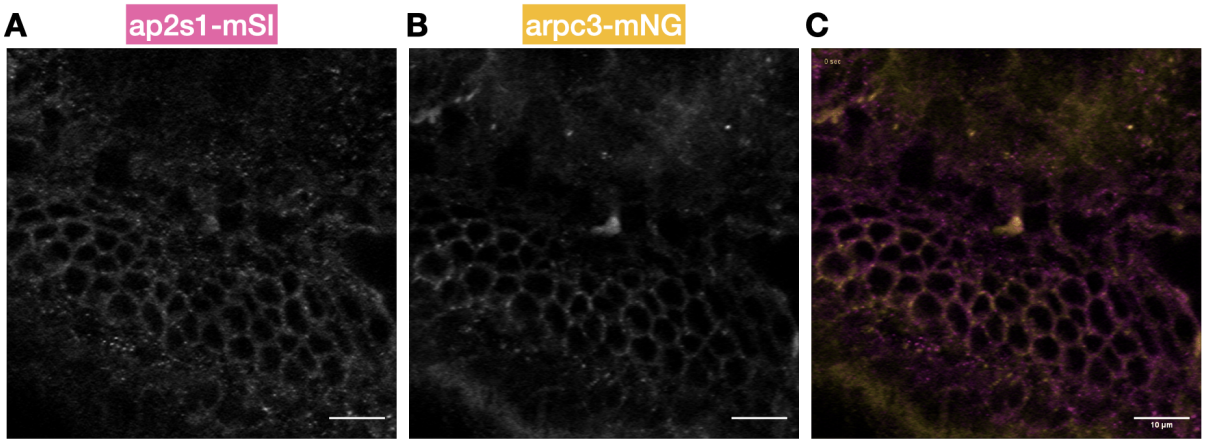


Figure 3.11 legend: AP2 and Arp2/3 labeled by ap2s1-mSI female (het) out-cross arpc3-mNG male (het), 72 hpf, A: ap2s1-mSI channel, B: arpc3-mNG channel, C: merge showing co-localization of CME structures with Arp2/3 *in vivo*

Additionally, I found Arp2/3 in developing erythrocytes (Figure 3.12 A and C) which did not overlap with dynamic membrane protrusions (Figure 3.12 A, labeled by an injected membrane-HaloTag construct), indicating that these structures are likely formin based[89]. Cells beneath endothelial cells (Figure 3.12 C), which are known to have strong Arp2/3 expression[90], that might be a part of the yolk syncytial layer also appear to have strong Arp2/3 enrichment.

Figure 3.12: Arp2/3 labeling in blood and endothelial cells

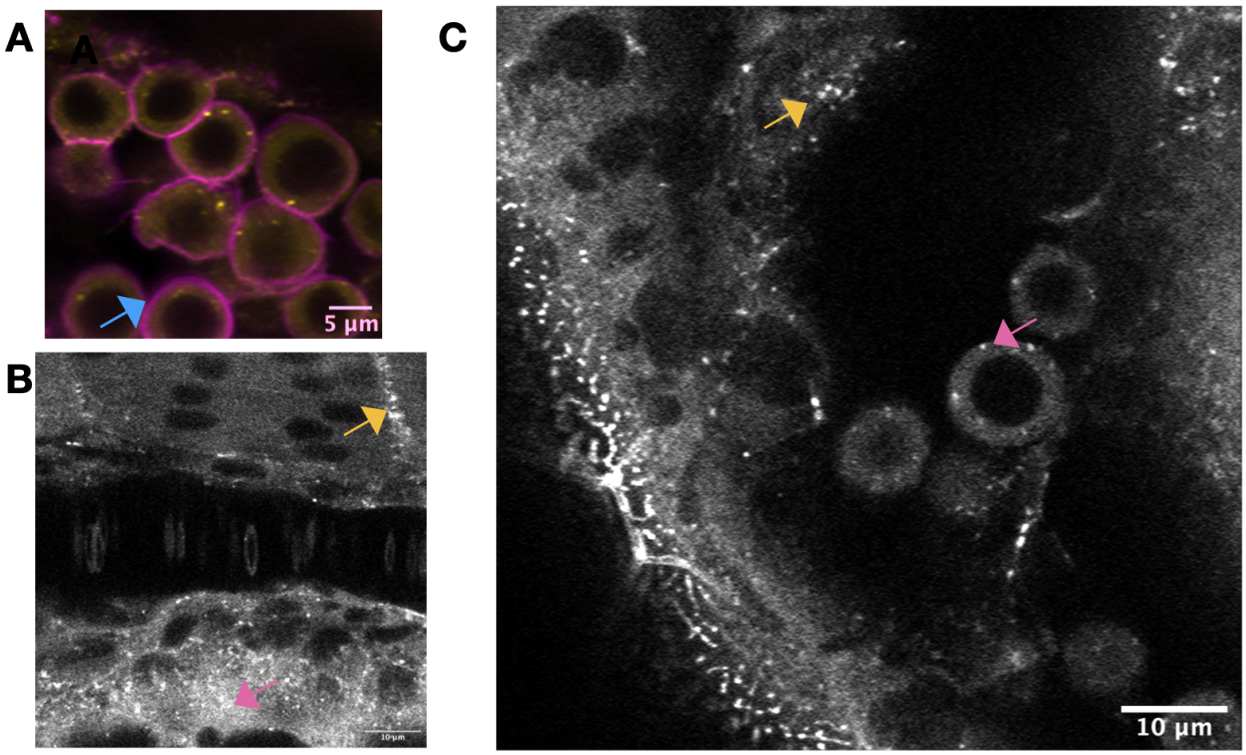


Figure 3.12 legend: Arp2/3 labeled via arpc3-mNG homozygous ix, A: 22 hpf, Arp2/3 puncta (yellow) found in primitive blood in intermediate cell mass with membrane-HaloTag construct labeled with Janelia Fluor 647 showing dynamic plasma membrane protrusions (blue arrow, possibly fillopodia) not labeled with Arp2/3, B: 30 hpf, common cardinal vein directly posterior of the Duct of Cuvier showing circulating blood and bright Arp2/3 structures near cells adjacent to yolk (magenta arrow, possibly yolk syncytial layer) and strong enrichment at cell-cell boundaries (yellow arrow) along anterior-posterior axis, C: 30 hpf, common cardinal vein flowing dorsal to ventral along left side of yolk, puncta in endothelial cells (yellow arrow) and circulating blood (magenta arrow)

The dynamics of transferrin internalization and CME in developing blood

By generating AP2 lines (ap2s1-mSI and ap2s1-mNG) and a clathrin line (clta-mNG) as well as a combination line (ap2s1-mSI, clta-mNG), I was able to visualize CME in developing embryos. There is some patterning of how the abundance of CME changes between adjacent cells in ear tissues such as hair cells and the endolymphatic sac (Figure 3.13 A). The same can be seen in neuromasts (Figure 3.13 B) which have a subset of cells in the center of the honeycomb structure that are brighter than neighboring cells. This result highlights that the demand for cargo or membrane remodeling is highly specialized between cells in the same organ.

Figure 3.13: AP2 labeling in ear and surrounding tissues

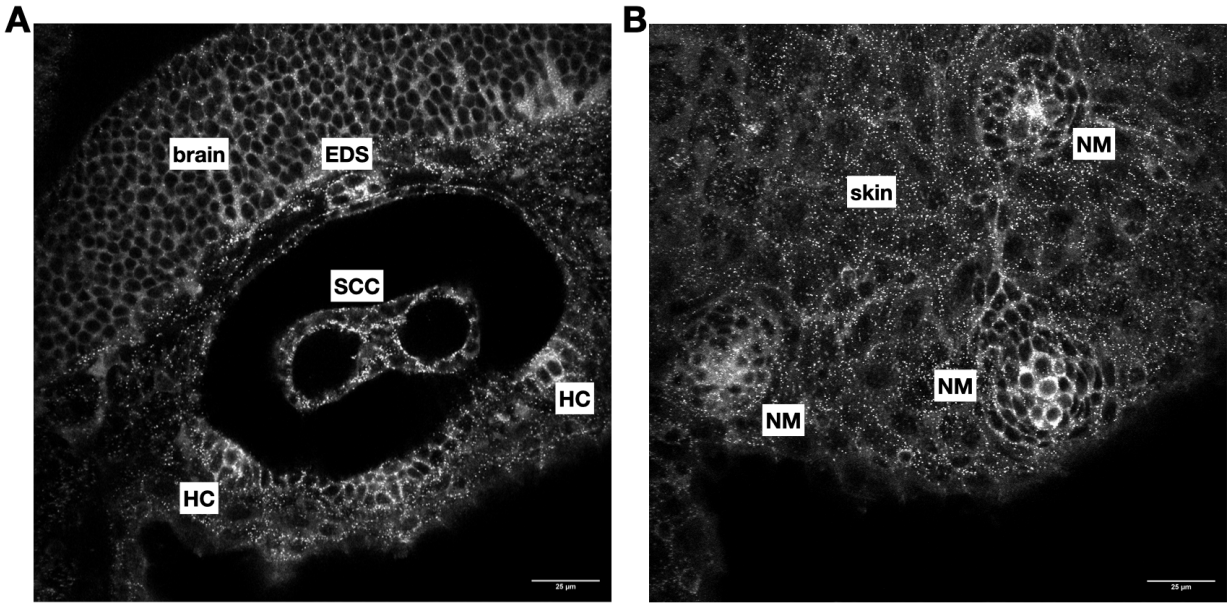


Figure 3.13 legend: AP2 labeling in ap2s1-mNG homozygous (including Casper) in-cross at 50 hpf A: endocytic vesicles organized into punctate structures in a subset of cells in hair cells (HC) bundles have dense AP2 labeling, brain, endolymphatic duct and sac (EDS), semi-circular canals (SCC), B: a subset of cells in center of neuromast (NM - honey-comb structure), basal surface of skin has dense labeling of endocytic vesicles

Cell specialization for CME is most pronounced in the caudal hematopoietic tissue[91] (CHT), a transient organ in zebrafish that houses hematopoietic stem and progenitor cells (HSPC) during their development (Figure 3.14). HSPCs give rise to all blood lineages in zebrafish. Cells in the CHT have the most dense CME labeling I have seen in the entire embryo, often so dense that individual puncta can often not be observed with Airyscan microscopy. To my surprise, when viewing the brightest structures in the CHT (Figure 3.14 C), I found that the ventral endothelial wall of the CHT has an apparent striped pattern of CME demand, reflecting some of the previously observed patterns of markers in the CHT[92]. I observed another example of enriched endocytosis in the posterior midgut (Figure 3.15) which is known to use CME for gut patterning[93]. These data highlight the ability to use genome-edited fish to study where a pathway is enriched and important for forming a tissue.

Figure 3.14: caudal hematopoietic tissue has dense and patterned CME usage

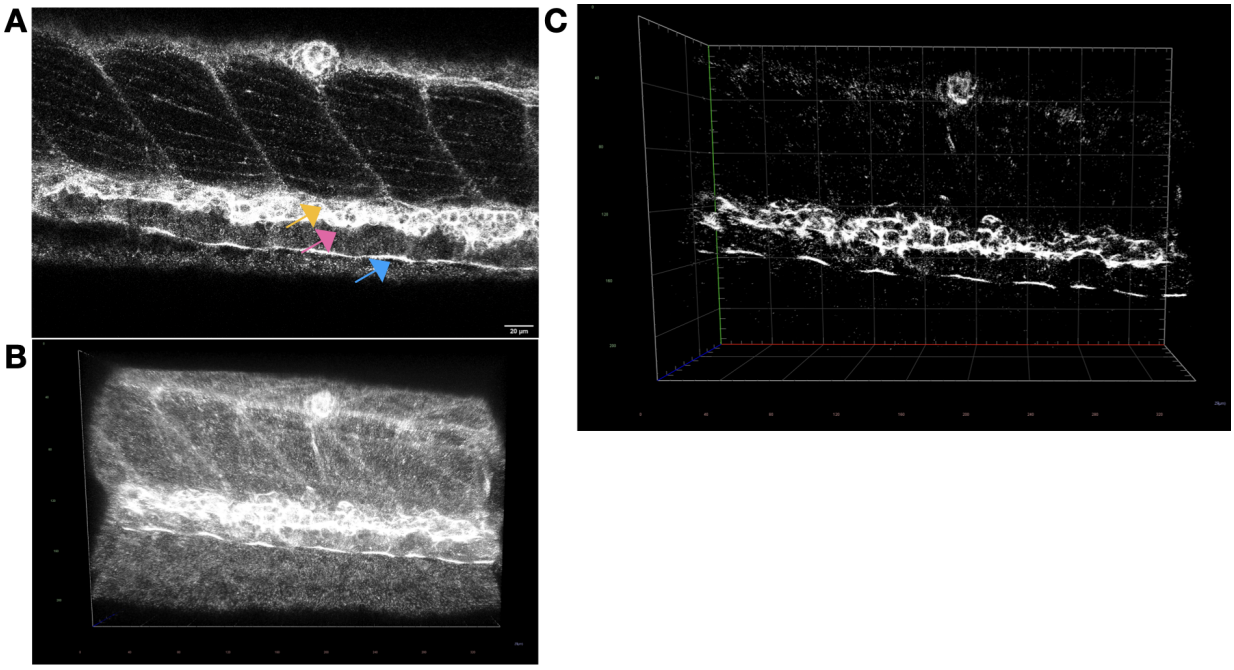

A
B
C
20 µm

Figure 3.14 legend: all data from single stack, AP2 labeled at 100 hpf via ap2s1-mNG homozygous, including Casper, in-cross: A: dense AP2 labeling in dorsal wall of posterior vein (yellow arrow) while circulating blood (magenta arrow) still has AP2 but devoid of clathrin, whereas polarization of CME density can be seen in ventral endothelial cells in posterior vein (blue arrow), B: 3-D rendering of volume imaged in A, highlighting stark contrast of caudal hematopoietic tissue compared to skeletal muscle and surrounding epidermis, C: clipped histogram of (B) cutting intensity values below ~99th percentile in brightness showing patterning of AP2 along ventral wall of posterior vein

Figure 3.15: clathrin's abundance spikes in the developing midgut

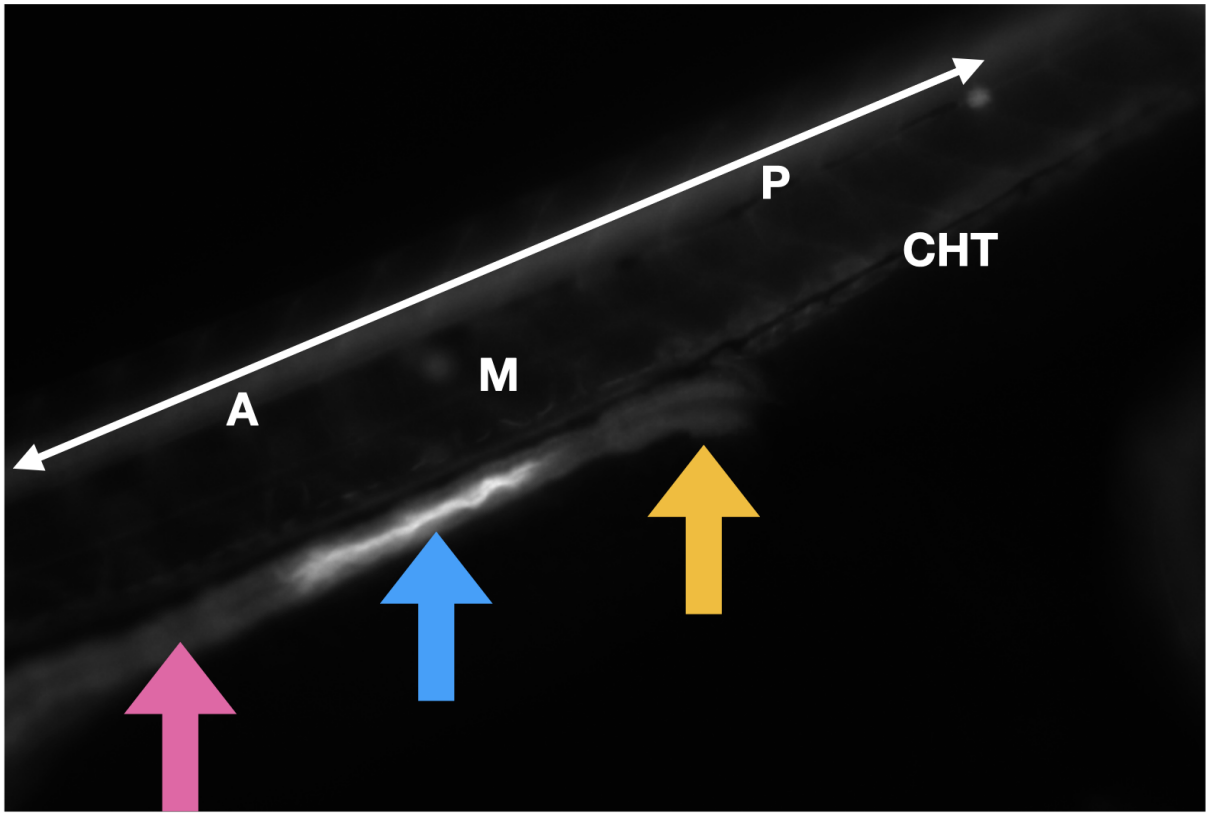


Figure 3.15 legend: Image of a 120 hpf Casper clta-mNG homozygous ix larva captured on a fluorescence stereomicroscope. Skeletal muscle (M) along the anterior (A) to posterior (P) axis shown in comparison to the midgut (blue arrow) situated between an anterior gut segment (magenta arrow) and the cloacae (yellow arrow).

One of the first experiments I did with my gene-edited ap2s1-mNG fish was to add already available cargos. Transferrin is an iron-carrying serum protein that has a major role in blood development by delivering iron to developing red blood cells (erythrocytes) to facilitate the production of hemoglobin[94]. Transferrin internalization is thought to be through CME of its receptors: the transferrin receptors. The zebrafish transferrin receptor was discovered through forward-genetics studies of the chianti mutant[95], a zebrafish model for studies of anemia. Whole-mount RNA in-situ hybridization was used to identify where the transferrin receptor transcript was localized, with the limitation that the imaging did not have the resolution to discriminate individual cells between the vascular endothelium and blood. The chianti paper claimed that one of the transferrin receptors, tf1a, is expressed in the CHT and blood, and additionally, a single-cell RNA-sequencing study showed that developing erythrocytes (up to 24 hpf) are enriched for transferrin receptors[90]. Therefore, to see transferrin activity at the protein level, I microinjected fluorescently labeled transferrin (Tfa-AlexaFluor647) into the developing CHT (Figure 3.16) and found a surprising result. At 22 hpf, I found that the majority of transferrin internalization took place in endothelial cells of the dorsal aorta and primary

vein around the CHT. There is an appreciated role for transferrin in endothelial cell spreading[96]. I could not see a clear transferrin signal in the developing erythrocytes (Figure 3.16 D). This data demonstrates that the chianti paper data is misleading because of its low resolution or because the transferrin receptor internalization activity is not proportionate to the abundance of the transcript which still might be in circulating blood. Single-cell transcript data shows that the erythrocyte population has higher transferrin receptor transcript abundance compared to the endothelial population at 24 hpf. My data suggests that erythrocytes are no longer appreciably consuming transferrin at high levels by 22 hpf. Additionally, I believe the key finding in this chapter is that endothelial cells are the population of cells that have the highest rates of transferrin CME activity, despite their relatively lower expression of the transferrin receptor when compared to erythrocytes. This data highlights the importance of studying pathways at the level of protein activity and that transcript abundance can be misleading in terms of whether there is high protein activity.

Figure 3.16: transferrin activity is most strong in endothelial cells, not erythrocytes, at 22 hpf

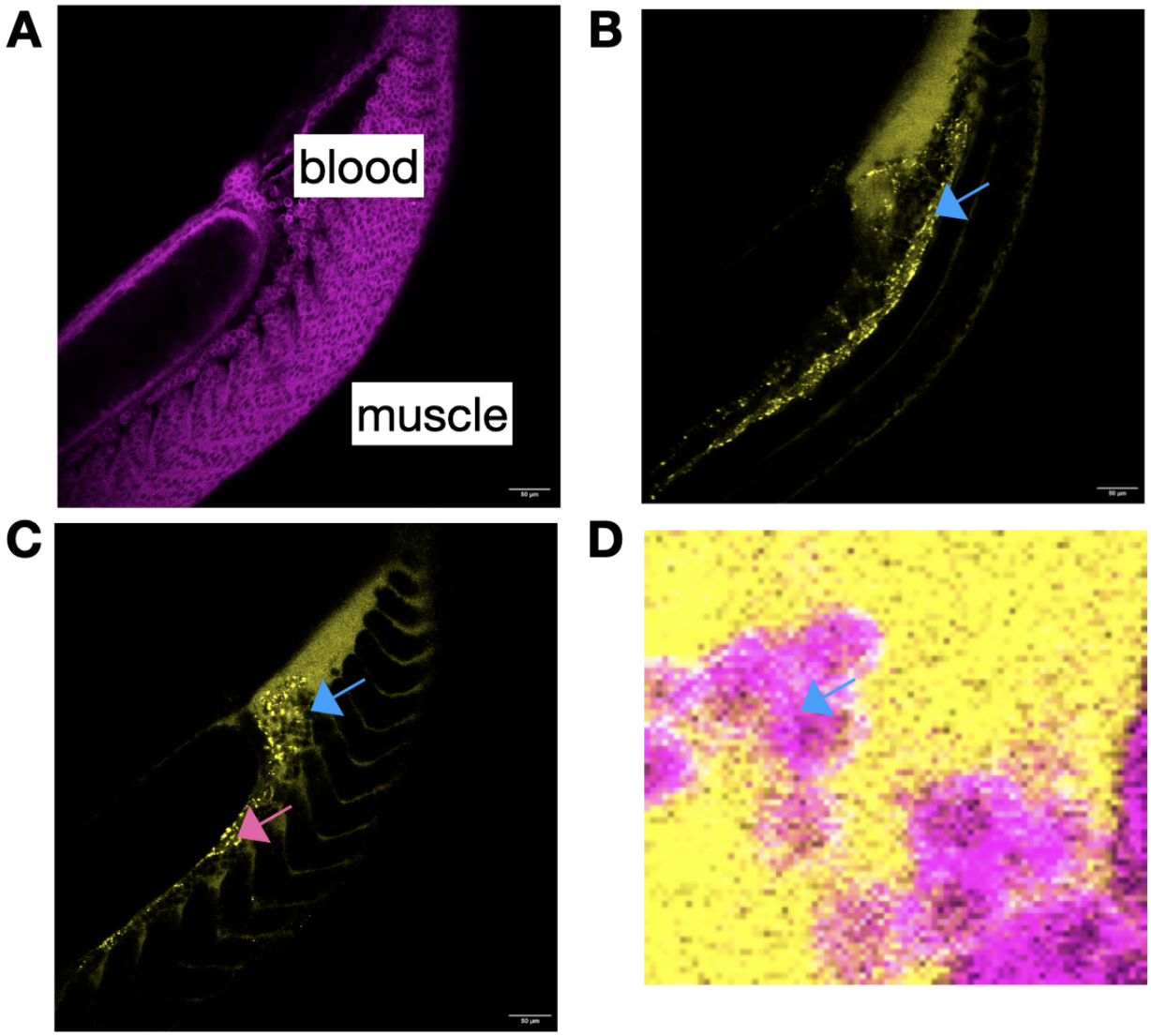


Figure 3.16 legend: AP2 labeled by ap2s1-mNG (heterozygous) at 22 hpf injected with transferrin-AlexFluor647 ~1 hour before imaging. A: anatomy of area imaged in 3-D stack, B: transferrin accumulation in dorsal aorta (blue arrow), C: transferrin accumulation in vein (magenta arrow) but not in primitive blood (blue arrow, enlargened in D with AP2). Note that the AP2 signal looks blurry in this micrograph because it is heterozygous and the fish was raised at a lower temperature to facilitate the 22 hpf injection; the localization of the AP2 fusion might be temperature sensitive.

It has been shown that HSPC have an upregulation of transferrin receptor expression during their specification[97]. Since erythrocytes are specified as early as 12 hpf[91], Erin, Emilia, and I found a way to inject developing embryos at 12 hpf by injecting through the yolk and into the neural tube, close to the site of the future intermediate cell mass. After delivering fluorescently labeled transferrin and waiting 16 hours until erythrocytes were

plentiful and populating the common cardinal vein (CCV), we found an enrichment of labeled transferrin in circulating cells (Figure 3.17 A). This experiment is tricky because erythrocytes are known to shed their organelles and this is apparent by imaging this same exact fish at a later time point (Figure 3.17 B) and observing that erythrocytes no longer have labeled transferrin inside. Further work is required to understand the ceiling stage that erythrocytes still consume transferrin. The difficulty with this experiment is the staging. I would slow down fish to inject by raising them near room temperature (22*C) overnight so they could be injected at ~12 hpf the following day, and would slow them down again to image at the onset of circulation ~28 hpf. Targeting other time-points of injection and imaging would require careful staging of parental crosses and temperature control of life stages to maintain any internalized transferrin before its eventual clearance. Further work is required to confirm that this ap2s1 fusion is low-temperature sensitive. This data supports the evidence that transferrin activity occurs early in the specification of erythrocytes.

Figure 3.17: iron internalization in developing erythrocyte occurs early in their specification

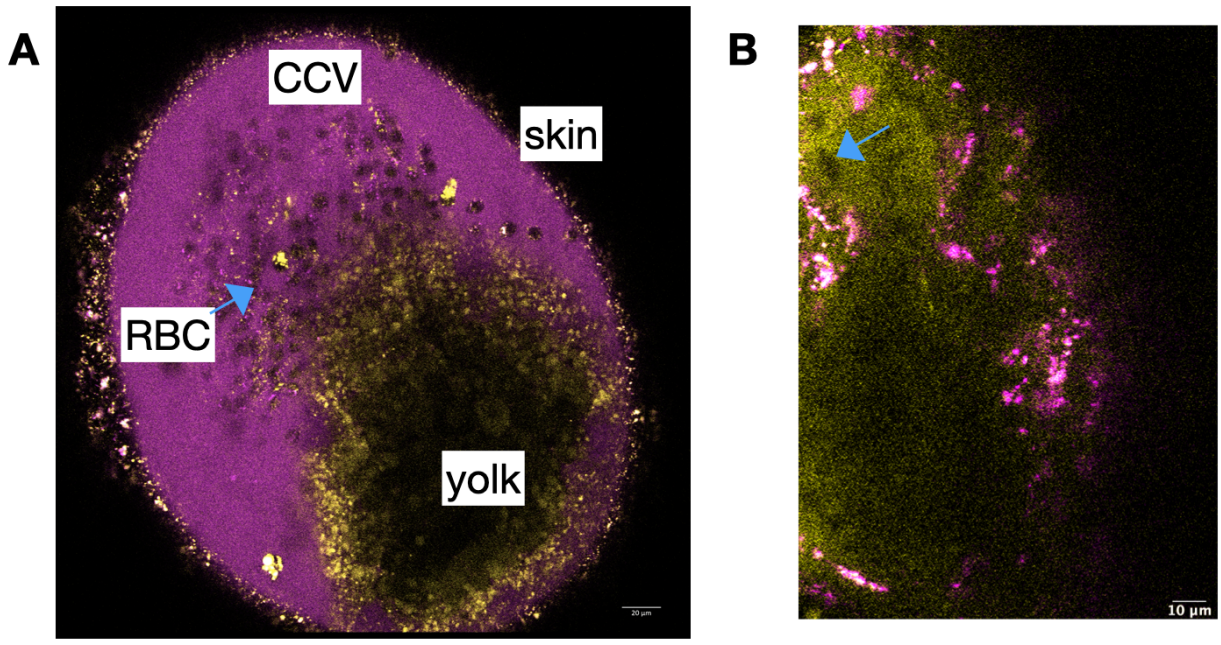


Figure 3.17 legend: same fish in both images, Casper fish injected at 12 hpf into neural tube with fluorescently labeled transferrin, yellow: transferrin-pHrodo, transferrin-AlexaFluor488, imaged at 28 hpf, A: cells adjacent to yolk in common cardinal vein (CCV) with erythrocytes (RBC: red blood cell, blue arrow) and other blood cells containing internalized transferrin, B: at 55 hpf, representative erythrocyte pooled in heart (blue arrow) showing lack of intracellular labeled transferrin

Although the idea that mature erythrocytes do not eat is not completely new[98], I generated further support that transferrin endocytosis is stopping during erythrocyte maturation by the observation that clathrin disappears in circulating erythrocytes by 72 hpf (Figure 3.18 B). The colocalization of clathrin and AP2 (Figure 3.18 A) at the 30 hpf time-point suggests that these maturing erythrocytes (still circular, not disk shaped at this point) may still be eating something orthogonal to transferrin and its receptor. I injected labeled transferrin at 30 hpf and observed an internalized population in the CHT endothelium but not circulating blood (Figure 3.19). The retention of AP2 in erythrocytes that have depleted clathrin could result from a perturbed and extended half-life of AP2 that does not degrade because of the tag. However, if this effect is real, it is possible that there is an alternative role for AP2 retention at the plasma membrane such as clustering receptors that maintain the erythrocytes net negative charge[99].

One other supporting piece of evidence for CME activity for iron internalization during early HSPC emergence is the observation that cells in the subaortic space (Figure 3.20: blue arrow) have elevated levels of clathrin and AP2. These cells are likely HSPCs that emerge from the dorsal aorta from the endothelial-to-hematopoietic transition[91]. This observation was hard to miss because of the dramatic downregulation of CME in the dorsal aorta after ~24 hpf. Recall that at 22 hpf (Figure 3.16), the dorsal aorta is still eating, but is later is polarized: the dorsal wall has little to no AP2/clathrin, but its ventral wall has a slightly higher density of CME in comparison to the strong maintenance of CME markers in the vein which was also present earlier at 22 hpf.

Figure 3.18: clathrin disappears while AP2 remains in circulating erythrocytes

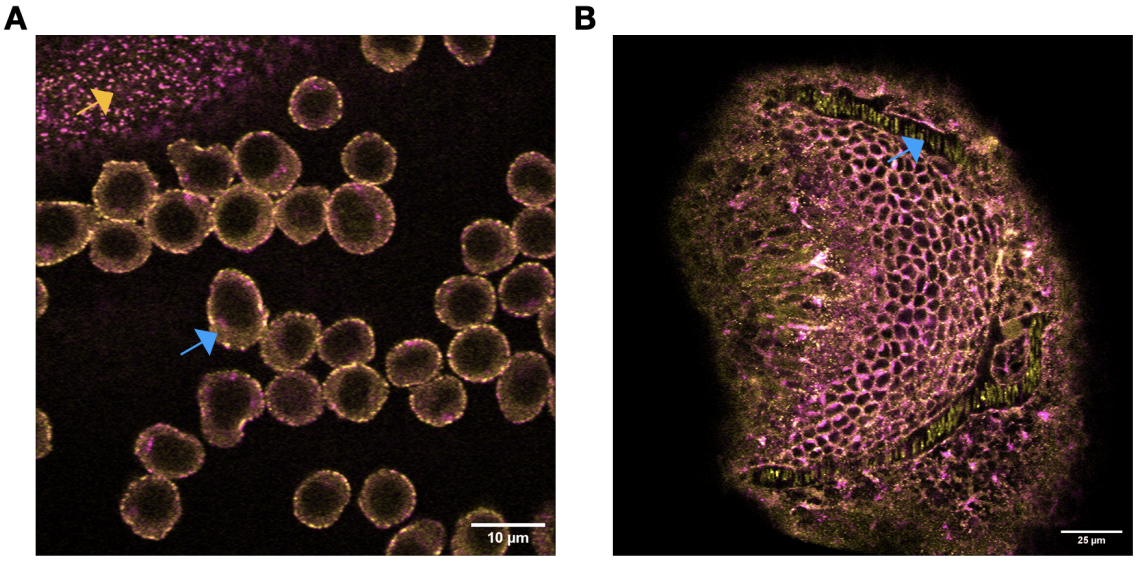


Figure 3.18 legend: clta-mNG (magenta) out-cross ap2s1-mSI (yellow) (both heterozygous), A: erythrocytes (blue arrow) in common cardinal vein displaying both

AP2 and clathrin adjacent to endothelial cells (yellow arrow), whereas B: has a loss of clathrin in circulating blood (blue arrow)

Figure 3.19: transferrin accumulates in caudal hematopoietic tissue, not circulating erythrocytes

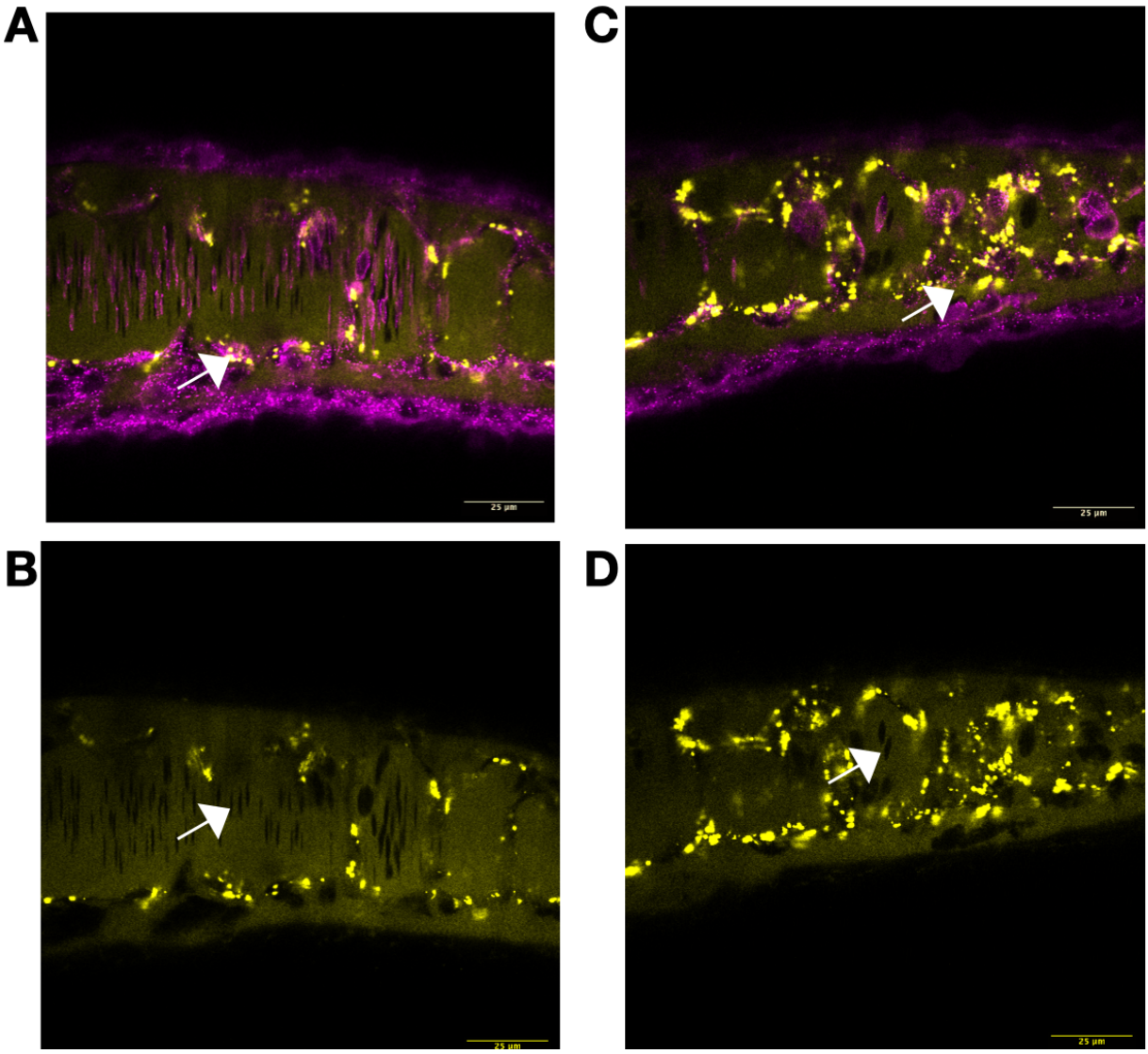


Figure 3.19 legend: AP2 labeled by homozygous ap2s1-mNG (including Casper) in-cross, transferrin labeled by AlexaFluor647 injected into circulation <5 minutes before imaging at 30 hpf. A: white arrows shows subset of endothelial cells in caudal hematopoietic tissue with strong transferrin accumulation. B: lack of circulating blood with internalized transferrin shown by black silhouettes (white arrow) devoid of dye in circulation shown in yellow background. C: 88 minutes later, transferrin accumulation is in endothelial cells but not visible in D: circulating blood visible by silhouettes (white arrow) within pool of circulating dye

Figure 3.20: polarity of CME in artery and veins, and spikes in CME during endothelial to hematopoietic transition

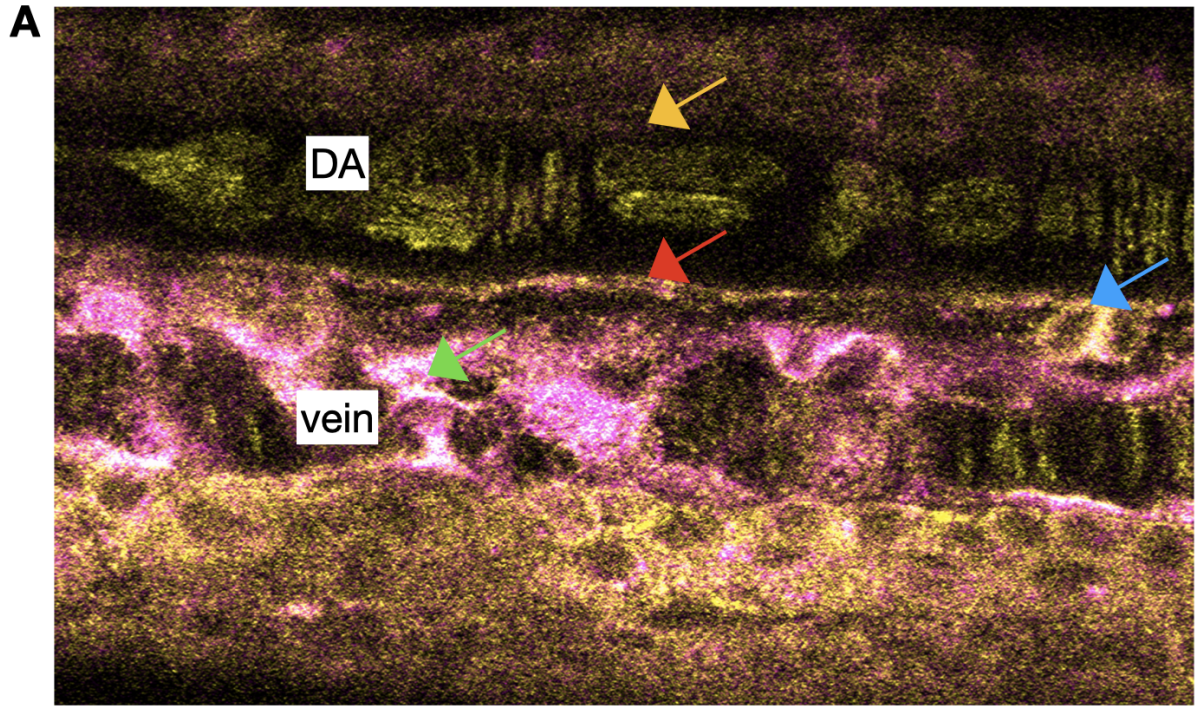


Figure 3.20 legend: heterozygous fish for both ap2s1-mSI and clta-mNG, homozygous Capser, in-cross'd, selected for zygotic double homozygous labeled at 80 hpf, A: dorsal aorta (DA) has polarity in CME density along dorsal (up, yellow arrow) and ventral (down, red arrow) axis with stronger presence on ventral surface whereas vein (green arrow) has high CME usage, ventral wall of DA undergoing endothelial to hematopoietic transition (blue arrow) has re-emergence of strong CME when hematopoietic stem and progenitor cells emerge

Discussion

When I initially set out to study CME *in vivo,* I thought I would spend much of my time making the same kinds of measurements that cell culture studies do: measure rates, lifetimes, and brightnesses. However, I have always had qualms with these measurements because they require a control that most studies are unwilling to do: a proper calibration of sensitivity. After all, what most people call "lifetimes" in CME markers is not the actual lifetime, but "measured lifetime". To measure real lifetimes, you need to know when the first few molecules of a pathway assemble (e.g. when AP2 lands at the plasma membrane) and know when the pathway stops (e.g. DNM2 bursts or puncta moves away from the plasma membrane). Knowing the intensity of a single fluorophore is in an optically complex substrate like a zebrafish tissue is not trivial. For instance, if I saw that the average lifetime of CME in a muscle cell is ~60 seconds and it

is ~40 seconds in brain, how do I know that the shorter brain events are not because I am unable to detect and track the dim events at the beginning and end of a CME trace because my sensitivity is lower in the brain? AO-LLSM can correct the aberrations that cause puncta to distort in shape, but there is no guarantee that the peak signal is fully recovered in corrected images. I do not think it makes much sense to compare intensities and lifetimes without proper calibration controls. Instead, I have uncovered trends that are more straightforward to interpret: the endothelium of the dorsal aorta and primary vein have strong CME to begin with (Figure 3.16), eventually the dorsal aorta loses its high density of CME (Figure 3.20) and becomes polarized along the dorsal and ventral walls while the vein maintains its activity. Meanwhile erythrocytes can eat transferrin early in their specification (Figure 3.16), and then are either not eating or eating something different by 30 hpf (Figure 3.18) and eventually lose the clathrin (clta at the minimum) machinery to eat at all. The disappearance of clathrin (clta) from circulating erythrocytes could be explained by the expression of clta's paralogue, cltb, instead of clta, and differences in splicing of clta that removes the last exon that is adjacent to moxNeonGreen. Furthermore, there could be underexplored pathways that repurpose AP2 to traffic at the plasma membrane without clathrin.

During my last semester, I was still eager to take up the challenge to measure lifetimes across developing embryos. Doing so required the knowledge of what the intensity of a single fluorophore looked like across various tissues that aberrate and degrade puncta quality. Fortunately, strategies in the Drubin Lab have adopted nanocage-based calibration standards that could be implemented in zebrafish[8,32,100]. The key innovation here relies on the expression of a set number of fluorophores assembled together in a single structure, the nanocage, that is diffraction-limited. Then, this puncta's intensity can be measured and related to the number of fluorophores e.g. 60 or 120 copies of a fluorophore. This has since been implemented in plants[101], and Meghan Turner (a graduate of Hernan Garcia's Lab - UCB, unpublished work only in thesis) implemented this strategy in flies which are grown at temperatures closer to zebrafish than cell culture (flies: ~room temperature, zebrafish: 28*C, cells: 37*C). I am grateful that Ian thought this was a worthwhile project and that he did all the cloning to make this construct for zebrafish expression while I was on the job market the past few months. Hopefully somebody gets to explore this soon and calibrate 60 or 120 copies of moxNeonGreen and subsequently back-calculate the fluorescence signal of one fluorophore so we can actually measure real lifetimes.

While I did not get to publish my various lines, I think they are all finding good use now. For instance, my clathrin and AP2 lines are now a cornerstone of studies in the Advanced Bioimaging Center that is focused on a new generation of adaptive optics technologies based on deep learning. I believe that is is important to pair these high

quality algorithms with reagents that are uniquely capable of showcasing their abilities: labeled organelles in a live organism that is not prone to the pitfalls of antibody labeling[102] and sample-to-sample variability from transgenic or over-expressed fusions.

I think it would be worthwhile to express the nanocages in zebrafish and measure how the intensity of 60 fluorophores changes as a function of depth regardless of the quality of the adaptive optics recovery of the point-spread function (PSF). For instance, if we were able to count “1000 arbitrary units” at the skin for one assembled nanocage, but measured “500 arbitrary units” deep into muscle with a perfect PSF, we would know that ~50% of our signal is attenuated by light traversing cells into and out of muscle. This would allow us to compare the abundance of a label at an organelle and compare this across multiple cell types.

It could also be useful to create various transgenic versions of the nanocage system. For instance, a standard reference channel (e.g. red-fluorophore tagged) nanocage can be crossed into lines of interest and the nanocage channel can be used to calibrate the adaptive optics regardless of the type of structure (e.g. non-punctate structures found in beta-catenin) found in the desired imaging channel. If healthy alleles can not be derived and maintained through adulthood, heat-shock or other inducible systems might be of help.

Early on in my beta-catenin project, I recommended we create ubiquitous transgenic alleles of the membrane-HaloTag and histone-HaloTag constructs. Fortunately, Samara was able to establish those alleles and we crossed them to my mNG-ctnnb1 line. I think there is a ripe opportunity to track nuclei during development and ask: which cells have green (i.e. beta-catenin co-localized) nuclei? Doing so may reveal the dynamics of canonical WNT signaling (CWS) during the development of an organ (e.g. ectodermal thickening of the dorsal otic vesicle during endolymphatic duct and sac emergence) or highlight cell types that have been underexplored in their CWS usage. These studies can reveal the nature of the CWS dynamics: for example, whether they are oscillatory, digitally switching, or terminally on. This mNG-ctnnb1 reagent is uniquely capable of answering such questions. To top this reagent off, Erin, Evan, Ian, and I cloned and prepared the reagents to tag the duplicated paralogue of ctnnb1, ctnnb2, which at first glance appeared to work in a low-throughput injection experiment.

One of the most rewarding aspects of this portion of my PhD was knowing I cracked gene-editing in zebrafish. Erin, Nate, Evan, and I were able to scale up our editing efforts and successfully target a half dozen other genes and see fluorescent F0s: paxillin (pxna), cadherins (cdh1 and cdh2), exocytosis (exoc2 and exoc4), dynamin (dnm2a), and endophilin (sh3gl1b). It has been a major headache to keep track of all

these alleles and exactly how they have been propagated to future generations. It is imperative to create modern solutions for this problem that allows researchers to unambiguously share their results with internal and external members. Zebrafish genetics is some of the messiest science I have ever had the pleasure of laying my eyes on.

To facilitate keeping track of the complicated family trees required to generate various allele combinations of these knock-in genes, Evan has been working on generating a database and program interface that contains a well-defined allele designation protocol and zebrafish facility inventory system. I have to thank Kevin Shin, my high school best friend and the best programmer I know, for giving Evan a ton of help when my expertise hit a wall and I was no longer able to give technical advice to Evan. The central goal of this project is to make a fully open-source, collaborative program that allows users to seamlessly keep track of hundreds of alleles, trace their lineages, get alerts on the status and health of fish (e.g. fish are old and need to be culled, fish are old enough to have sex), and aid users in designing complicated combinations of desired alleles (e.g. cross this fish with that fish then cross that fish with this fish, while the program keeps track of their ages and sexual availability, too). I think Evan is in the right place at the right time to get this project across the finish line in beta in the next few months.

Methods

Animal work

All zebrafish work was done in accordance with Berkeley AUP 2020-10-13737.

Genome Editing

Fluorophore (mNeonGreen and mScarlet-I) and linker (3xGGS) coding sequences flanked by homology arms (Table 3.1) for each gene were cloned into a plasmid backbone (Table 3.2) using isothermal assembly. Homology arms and the fluorophore were amplified off the plasmid starting and ending at biotinylated-oligonucleotide landing sites, and purified amplicons with 5’ biotin moiety (25 pg) were injected into 1-cell embryos of AB in-crosses with Cas9 mRNA (50 pg, TriLink BioTechnologies) and sgRNA (5uM, IDT, Table 3.3). Founders were raised and out-crossed to ABs or Caspers to establish F1s which were individually Sanger sequenced and screened for proper integration with junction amplicons.

Table 3.1: homology arms

Protein Functional Fusion	LHA + linker + (mSI or mNG) + RHA (5' to 3')
ap2s1-mNG	tgctaaagttgactgttgagttttacagtgtagttagtgttttacactatcccaaaataagtctgatcatt gttttactaaggcatgtgaaatatttattatcgactaaaatattgttattgttgtcaatatgtgagctcatg tttagtatgccattcattttgtcaaaaataaatgaataaaaaaaaacatgtttagagagtcagcaag aatatactatgaaacttattagaaggcagttaaaaagatcaattcaatatataaaagcaaaaaca tatacggttgaagtcagaattattagccccctttgaatttttttttctttttttaaatatttcccaaatgatgttt aacagagcaaggaaactttcacaatatgtctgattatatttttttcttctggaaaaggtctttcttgttttatt tcggctagaataaaagccatttttaatttttaaacaccattttaaggacaaaattattagccccttaaa ggtaatttttttttttttacgatagtctacagaaccattgttatacaataacttggctaattaccctaacct gcctagttaacctaattaacatacacgcatacatacctttttaacacgcatacaaacacacacaatttt ttattttcacaaaataatgtttttgaaattcccagagatgggttgcggccggaagggcatctggtgtgt aaaaatgtgctggataagttggcggttcattccgctgtggcgaccccagataaataaagggacta agccgacaagaaaattaatgaatgaacgaatgaatgtttttgaagtttaagtttgtgtagaactgcg ttatagtcaagaaatatcccaaaaattgagtcatgttgacaaacatacattgattaagttaacttaatt gtttttgcaaattaaagtggattggaactaaagcaattaatttttctctcaaaaaacttaagaattgatt tagctcaactaacaaaaatatttttgagtgcaaacacatgtaaaaaatgcacgaagcagcaagtt tttagttgaaaacagcggccctatatgtgaactgctattaagaagtattgtggatgtgtttcaggtgta tacggtggtggacgagatgtttctagcaggagagatcagagaaaccagccagacgaaggtgct caagcagctcctcatgctgcagtccctggaggGCGGatccggtggatccggtggatctgtTag CaagggcgaggagGACAAtATGGCCTCtCTGCCCGCAACACACGAGCT GCATATTTTCGGAAGCATCAAcGGcGTGGATTTCGAtATGGTtGGgC AaGGaACTGGAAACCCAAATGAcGGaTACGAGGAACTGAATCTGAA GTCAACCAAAGGCGACCTCCAaTTCtcaCCTTGGATTCTcGTtCCCC AtATTGGCTATGGaTTTCATCAaTATCTGCCaTAtCCTGAtGGAATGTC ACCATTTCAaGCcGCTATGGTGGATGGATCTGGCTACCAaGTCCAC cgcACCATGCAaTTTGAGGACGGcGCCtccCTGACTGTGAACTACCG CTATACCTACGAGGGATCtCATATCAAGGGCGAAGCACAaGTtAAAG GaACAGGATTCCCAGCTGAcGGCCCCGTCATGACAAACTCTCTGA CCGCCGCCGACTGGAGCCGGTCCAAGAAAACTTACCCTAACGATA AGACCATCATCTCTACCTTCAAaTGGAGTTATACCACcGGCAACGG aAAGCGcTACAGAAGCACAGCCCGAACTACCTATACTTTTGCTAAG CCtATGGCTGCAAACTATCTGAAAAATCAGCCTATGTAtGTcTTCcGA AAaACcGAattgAAGCACTCCAAAACAGAACTGAATTTCAAGGAgTG GCAGAAGGCTTTTACCGATGTtATGGGcatggacgagctgtacaaAtaaactgt caatcagtcaaccaaccaatcaaacaaccaccattgtgtcaggctccgcccaccccgctgcag ccaagccatatgatttgtgtgtatgagtgttgtcatgctgaaaatcatcctctaacatttcacctttttcttt acactcgcatccacctcacaggatgttgccgttctgtcctgcaacttcatgaattttcgaggcagata agaattcgtaaaacttcctcggttatgtgacctgattatatcttggcatagccagaattgaattcttca gcgattttatagcacactagacttcagtgtttccgactggaagcctctccaagcctgttcattaacact

	aacatagataatagcgtcatgtgggccaaaaaactgccatggcacgaaagtgacagtcataatt gcaagattccggggctgcactgcaatatcgcaatgagaaaaacttgatttgatgtccatttaagga aattcccatatatgtactattaaaaaaagaatcaaatgtttaatatgctgcactgcaaataataaattt gttactgctttgtcttgttttctggtacaagtatccatattgttaaaactccaaatacgtaaaggtgatctt cctgttggtgactatatctttttttatacaattaagatgttcataatggtgtcacaaaaagggtttcgcaa caactgttggcacttttgagagtaaaaaaaaaacatacaggcaaaaccaaattaactgaccgttt cagtatggagatgtcattaatccatgaacagttcctgcttgtcttaactatttctgaattgtaagaggc gattcggttgtaactaatgttaacatggctgctaaaaaaagtgtcaatatggacctttctggattctaa taagatttggaaatgaagatataaaacagcaaatacataaataccattgttattatttacatccctca aaatggtctacccactactgattcggtgaggtcctaagaagtgaaattaaaatttatgtcaaacaa aatacattattttagatccacagctttggccaattgtattctttttgttttgcctattattattatttttttttttttca ctctaaaaagtaataatcgggttaatcagcctaatggtgaatttaaagaaaaatgaaggagaaa acaaaatatttccttgccatactcttac
ap2s1-mSI	tgctaaagttgactgttgagttttacagtgtagttagtgttttacactatcccaaaataagtctgatcatt gttttactaaggcatgtgaaatatttattatcgactaaaatattgttattgttgtcaatatgtgagctcatg tttagtatgccattcattttgtcaaaaataaatgaataaaaaaaaaacatgtttagagagtcagcaag aatatactatgaaacttattagaaggcagttaaaaagatcaattcaatatataaaagcaaaaaca tatacggttgaagtcagaattattagccccctttgaatttttttttctttttttaaatatttcccaaatgatgttt aacagagcaaggaaactttcacaatatgtctgattatatttttttcttctggaaaaggtctttcttgttttatt tcggctagaataaaagccatttttaatttttaaacaccattttaaggacaaaattattagcccttaaa ggtaatttttttttttttacgatagtctacagaaccattgttatacaataacttggctaattaccctaacct gcctagttaacctaattaacatacacgcatacatacctttaacacgcatacaaacacacacaattt ttattttcacaaaataatgtttttgaaattcccagagatgggttgcggccggaagggcatctggtgtgt aaaaatgtgctggataagttggcggttcattccgctgtggcgaccccagataaataaagggacta agccgacaagaaaattaatgaatgaacgaatgaatgtttttgaagtttaagtttgtgtagaactgcg ttatagtcaagaaatatcccaaaaattgagtcatgttgacaaacatacattgattaagttaacttaatt gtttttgcaaattaaagtggattggaactaaagcaattaatttttctctcaaaaaacttaagaattgatt tagctcaactaacaaaaatatttttgagtgcaaacacatgtaaaaaatgcacgaagcagcaagtt tttagttgaaaacagcggccctatatgtgaactgctattaagaagtattgtggatgtgtttcaggtgta tacggtggtggacgagatgtttctagcaggagagatcagagaaaccagccagacgaaggtgct caagcagctcctcatgctgcagtccctggaggGGatccggtggatccggtggatctgttagc aagggagaggcagtgATTAAGGAATTTATGAGATTTAAGgtccacatggAAGG GTCCATGAACGGACACGAATTTGAAATAGAGGGGGAAGGGGAAG GTCGACCATATGAAGGGACGCAAACGGCAAAGTTGAAAGTTACTA AGGGTGGACCGCTCCCCTTTTCCTGGGACATATTGTCTCCGCAAT TTATGTACGGAAgcagggctttTATTAAACATCCGGCGGATATACCAGAC TACTACAAACAGtcattccctgAGGGGTTCAAATGGGAAAGAGTTATGA ATTTTGAGGACGgaggggcagtGACAGTCACGCAAGATACGTCATTGG AGGATGgaacactgatTTACAAAgtcaagttgcGGGGCACAAATTTTCCACC GGATGGGCCAGTAATGCAGAAAAAAACCATGGGATGGGAGGCGT CCACAGAGCGGCTCTACCCCGAGGACGGTGTCCTTAAGGgagatata aaAATGGCTTTGCGGTTGAAAGACGGCGGCCGTTATCTGGCGGAT TTCAAAACGACCTACAAAGCGAAGAAGCCTGTTcagatgcctggagcctac aaTGTCGATCGGAAACTCGATATAACTAGCCATAACGAAGACTACAC TGTTGTCGAACAatacgaacgaTCAGAAGGACGACACTCTACTGGGG

	gaatggatgaACTTTACAAAtaaactgtcaatcagtcaaccaaccaatcaaacaaccacc attgtgtcaggctccgcccaccccgctgcagccaagccatatgatttgtgtgtatgagtgttgtcatg ctgaaaatcatcctctaacatttcacctttttctttacactcgcatccacctcacaggatgttgccgttct gtcctgcaacttcatgaattttcgaggcagataagaattcgtaaaacttcctcggttatgtgacctgat tatatcttggcatagccagaattgaattcttcagcgattttatagcacactagacttcagtgtttccgac tggaagcctctccaagcctgttcattaacactaacatagataatagcgtcatgtgggccaaaaaa ctgccatggcacgaaagtgacagtcataattgcaagattccggggctgcactgcaatatcgcaat gagaaaaacttgatttgatgtccatttaaggaaattcccatatatgtactattaaaaaaagaatcaa atgtttaatatgctgcactgcaaataataaatttgttactgctttgtcttgttttctggtacaagtatccata ttgttaaaactccaaatacgtaaaggtgatcttcctgttggtgactatatctttttttatacaattaagatg ttcataatggtgtcacaaaaagggtttcgcaacaactgttggcacttttgagagtaaaaaaaaaac atacaggcaaaaccaaattaactgaccgtttcagtatggagatgtcattaatccatgaacagttcct gcttgtcttaactatttctgaattgtaagaggcgattcggttgtaactaatgttaacatggctgctaaaa aaagtgtcaatatggacctttctggattctaataagatttggaaatgaagatataaaacagcaaat acataaataccattgttattatttacatccctcaaaatggtctacccactactgattcggtgaggtcct aagaagtgaaattaaaatttatgtcaaacaaaatacattattttagatccacagctttggccaattgt attctttttgttttgcctattattattatttttttttttttcactctaaaaagtaataatcgggttaatcagcctaat ggtgaatttaaagaaaaatgaaggagaaaacaaaatatttccttgccatactcttac
arpc3-mNG	ctaaaggatttttttgctgttttcttgcattgggacgtttttttaaatgacagtgaaggaaaaatgtgcttc aaagtagttactgtaatgcaaacaagaagaaccgagggaaaaaaatcaaggaaagattgctt aaggcacatttttttttacttttttaaggtcacattccttttcatgatagttaaaacatttgtctacaagtgttttt tttaatgacaattaaaccaatttttctttttcagattatactgtaaaatagaacttctgaacatttatatttgt gttgctctcaaacgaactctctgccacaaactgaatattatttacaactttattacctgtttttcacagcc tatgcaggttctgttcactaaagtagacgtgcgcgcgtctatcatttaagtagcaatgaggtaacag ctcacggtcagcagctcacgtcacttgtgatttcccggccgcaaatacagtattatatgaagacaa aaatgacatttttaggtgaattgtaccttataagtacagtatgtgactcttttaatatcatttggaggtgt ccatgtcgctgagcaacccatgtgaaacaatgaaaagactcgtgcagccgcctttgcttctttttctg aacttgaaaatatgacagcagcagttggcagaaacatagaaatgttggattaagatcgtctagcg ggtcaaatcgagatcgcaatctttttttcaattaattgtgtattcagctattattttctcaatcacatttctaa agtcattatTataatgttcaacctgtatgtttgttcTagaaagttttgcattatatggaaacattttaaata aaattGaaacattttattaaattgtactacattgtacactgtagaatatatgttccctaaaatactcac accaaaattcacaaaaacattttgacaaacagtgattatttaataacactttatttttttttatctacagt ggtgggtgtgttttgtaaagaaacagttcatgaacaaaagtctgtctgcaccgggccaggcGG atccggtggatccggtggatctgtTagCaagggcgaggagGACAAtATGGCCTCtCT GCCCGCAACACACGAGCTGCATATTTTCGGAAGCATCAAcGGcGT GGATTTCGAtATGGTtGGgCAaGGaACTGGAAACCCAAATGAcGGaT ACGAGGAACTGAATCTGAAGTCAACCAAAGGCGACCTCCAaTTCtc aCCTTGGATTCTcGTtCCCCAtATTGGCTATGGaTTTCATCAaTATCTG CCaTAtCCTGAtGGAATGTCACCATTTCAaGCcGCTATGGTGGATGG ATCTGGCTACCAaGTCCACcgcACCATGCAaTTTGAGGACGGcGCCt ccCTGACTGTGAACTACCGCTATACCTACGAGGGATCtCATATCAAG GGCGAAGCACAaGTtAAAGGaACAGGATTCCCAGCTGAcGGCCCC GTCATGACAAACTCTCTGACCGCCGCCGACTGGAGCCGGTCCAA GAAAACTTACCCTAACGATAAGACCATCATCTCTACCTTCAAaTGGA GTTATACCACcGGCAACGGaAAGCGcTACAGAAGCACAGCCCGAA

	CTACCTATACTTTTGCTAAGCCtATGGCTGCAAACTATCTGAAAAATC AGCCTATGTAtGTcTTCcGAAAaACcGAattgAAGCACTCCAAAACAGA ACTGAATTTCAAGGAgTGGCAGAAGGCTTTTACCGATGTtATGGGca tggacgagctgtacaaA
cav1-m NG	ccactaaactgacaaaacgcaaaatagttacgtttccttgtgagatcaggctggtttttcggtaacg gttttgaccatatgattgattcaaaggtcttagaatGcagtAagtgggttaattggaaaacagttctg caaaatgaaaTcatgtctgatttccatcatacaatctaaaacgatctaattccgaaaaatctttttata cctccaggtggactttgaggacgtgatcgccgagcctgccggcacctacagcttcgacggcgtgt ggaaggcgagcttcaccaccttcacagtaaccaaatattggtgctacaggctgctgacagcgct ggtgggcatcccactcgccctAgtatggggcatcttcttcgccatcctctccttcatccacatctggg ccgtggtgccttgcgtgaagagctacctaatcgagatccactgcatcagtcgagtttactccatctg CgtgcacaccttctgcgacccactctttgaggccatggggaaatgctttagCaacgtccgggtca ctgctactaaggttgtggcgGatccggtggatccggtggatctgtTagCaagggcgaggag GACAAtATGGCCTCtCTGCCCGCAACACACGAGCTGCATATTTTCG GAAGCATCAAcGGcGTGGATTTCGAtATGGTtGGgCAaGGaACTGGA AACCCAAATGAcGGaTACGAGGAACTGAATCTGAAGTCAACCAAAG GCGACCTCCAaTTCtcaCCTTGGATTCTcGTtCCCCAtATTGGCTATG GaTTTCATCAaTATCTGCCaTAtCCTGAtGGAATGTCACCATTTCAaG CcGCTATGGTGGATGGATCTGGCTACCAaGTCCACcgcACCATGCAa TTTGAGGACGGcGCCtccCTGACTGTGAACTACCGCTATACCTACGA GGGATCtCATATCAAGGGCGAAGCACAaGTtAAAGGaACAGGATTC CCAGCTGAcGGCCCCGTCATGACAAACTCTCTGACCGCCGCCGA CTGGAGCCGGTCCAAGAAAACTTACCCTAACGATAAGACCATCATC TCTACCTTCAAaTGGAGTTATACCACcGGCAACGGaAAGCGcTACA GAAGCACAGCCCGAACTACCTATACTTTTGCTAAGCCtATGGCTGC AAACTATCTGAAAAATCAGCCTATGTAtGTcTTCcGAAAaACcGAattgA AGCACTCCAAAACAGAACTGAATTTCAAGGAgTGGCAGAAGGCTT TTACCGATGTtATGGGcatggacgagctgtacaaAtaaggacagagtgaaagagaa gagccagagggatcgggtgggatgggggaatggaaaccaatgtccaatgttgggaggtaaat caaacgtagcctctgtaataaagaggaaacaaaaTatgtgagcagaagtgaagggttgactct ccattcctgttcaaatgcatctgtgagatttgtctactcagctttggatttaggggacgtgtgttgaaaa tctctaattttctttgtcatttgttagcatttatcattgtaggaatactgtttcagccttacttttattgtgacag ttcatggaaaaatgctaattctGttaagtgtttactgacactaatgtcattccaaattttccaagctcct aaagtgcaataatagtggtttgtatgactagtgtgatataagaacctactcaggatctcagttggtac ctctgtgcgggatgtggatggatatttttttttgtctgtctgttttctttaatgtttatgcaaacctgtttaccctt attcagattttgtactggttatatcaagtgtatttgatttctgttctgtcatgtataacagcagatgatcaa ataaaagtctattgaagagaaacagtattattattagcatctaaagacacgtcagttttctgatttagc ttttaatattatattttcagtattgtagccatacagtatgtgggtcaaagatgtaaaaagcttttatgagtt tctatttcagttggtttccaagacaacattttctcatctgatatacaagacaaatttatttatgtagcaca tttcatacacaatggtaattcaaagtacttcacataaccaggaataaaagaaacaggtataagaa aataaaaacaacaaagaataaaaatagattaaaatgtgttgaaagtctcaaacgaaagaaaa agaaaagaaagacacaatagtgtgatctgttgaacgaagcacagggctcattcagtaaaggca cagctaaactgatgtgttttgagtcttgagtagaacgtgcctaatgttggagcacatctgatcatttct ggaagctgatttcagcagctaaag

clta-mNG	ggcttgccagaaacttggtttagttcccctcccccatgctcagcagaaggccccgcttttggtcact gccaaacgcatggcacgttcagcctttgcgcaaagatttagtggttaaaataagccaatgggcct gtcaatcgatacattgaaggctcattagcgtacgccgttcTCCGTCctctgaccctcaagaga cgaAgaacatcccttaattcaaagtgcgaccaattgttttaaccgactgaatcgctgcaaaacttttt aattatggctgttctggaaacctttttcctgccttcaagatAgagactagttctggtaatggtgggttgg gggcatgatgctcaaactaccatctgttttggatgtggtcacttgaagaatgtctctgaatgtatcgat ggatggaaaataagcacctggacgctcctgcttttggctcacgcacctctaacctcctaaccaaa aagctcctggtgaaacatttcagattttagacaactccttactgagttctttggaaattgatTgtttgaa gttctggaggctttgaaggaggtcagcggtgctgcgctcaagcaggacgggtctgcgagtgttcc ccgtgtctatctgaagctgattttggtttgtcatcggcagGGCGGCTGAGGAAGCCATG GTGTCGGAGCTGGATGAAAACAGTCCTGGCACAGAATGGGAACG TGTGGCGCGTCTTTGCGATTTCAACCCTAAATCCAGCAAGCAGGC AAAGGATGTGTCCCGCATGCGTTCAGTGCTCATCTCTCTTAAACAG GCTCCGCTCGTCCGCggcGGatccggtggatccggtggatctgtTagCaagggcg aggagGACAAtATGGCCTCtCTGCCCGCAACACACGAGCTGCATATT TTCGGAAGCATCAAcGGcGTGGATTTCGAtATGGTtGGgCAaGGaAC TGGAAACCCAAATGAcGGaTACGAGGAACTGAATCTGAAGTCAAC CAAAGGCGACCTCCAaTTCtcaCCTTGGATTCTcGTtCCCCAtATTGG CTATGGaTTTCATCAaTATCTGCCaTAtCCTGAtGGAATGTCACCATTT CAaGCcGCTATGGTGGATGGATCTGGCTACCAaGTCCACcgcACCAT GCAaTTTGAGGACGGcGCCtccCTGACTGTGAACTACCGCTATACCT ACGAGGGATCtCATATCAAGGGCGAAGCACAaGTtAAAGGaACAGG ATTCCCAGCTGAcGGCCCCGTCATGACAAACTCTCTGACCGCCGC CGACTGGAGCCGGTCCAAGAAAACTTACCCTAACGATAAGACCAT CATCTCTACCTTCAAaTGGAGTTATACCACcGGCAACGGaAAGCGc TACAGAAGCACAGCCCGAACTACCTATACTTTTGCTAAGCCtATGG CTGCAAACTATCTGAAAAATCAGCCTATGTAtGTcTTCcGAAAaACcG AattgAAGCACTCCAAAACAGAACTGAATTTCAAGGAgTGGCAGAAG GCTTTTACCGATGTtATGGGcatggacgagctgtacaaAACAGCTTCAAAG TAAACGAAAATCCAACAATCCCAGCAACCATTGGGTTTCACAACAT CGATTTCCTTTGTGGCCCTATCCCAATGGCTTTGAGCTCCATATATG AATAATCTATATTATATTATATGTATGTGTAATATGATATAGTTATGTCTA TAAATATTTCATCTATAAATATATACATGCATACATAAATATATATAACAA TGTTTTTATTTTTGGTGTCTATTTTATGGACCAAACTCCATGTTCCAC CTATTTTTCCCCTCGCCGTGTAGAGCCAGCGTGTGGCATTGTGAC ACTATCGATCTATGCATTTCCCTCGTGAATCTGTTGGGGAATCCAG TACTaACGTCTTGTGTATGTTGTATACAGATGGTGTGCGTGTGTGTG ACTcTACATAAAGACTGGTCTTCTACACTTCATTGTGAGATTATGGC ACAAGTTGTTGTTGTTGTTAAACAGAGCTGTACCCCGGTAAACGTT AACTCATGTTGTCCATTCTGTTTCAGTTTGATTAAACACATTTGAAA TGGTAAgtgtttgtgtgtcgccaccattttatttgcactgtttggaggaaaAcgtgagatctgcttttt ctttaagttttcttaaaaatcctttacttaaagtgaagaaatgtaggtttaattttaaagttgctgtatata cattttgactcGTctaaagcatcaaaataccataatatgtttgcagatatttgagaaacatgccaa gtaaatatacttgtttatctgaaaaacCatgctgaagtcagatattctgcttttataaatgtGcattac gtgccaaaacgactgtctCtgttttgggttttttaacctgcacattgccagtttagccaattataattca gcaccccgggttgcctcggtggaaaaccaaatatttcattcattcagtcaggaaggctctcagag

	catgtgcctgtgaccaaaagtctggtggacagtagcagactcctaaatgagacgcagattcaaagttccacatg
mNG-ctnnb1	tgactgtgttaagaagcctccctgcttggaggaggagatggatggttggagcgatgctaatagctcaagctttatcatcgtagttctcatggaatgatcccaggttgttgttttttttctggctagaaaaacattgattcatgaggctttttaaaatcatgatttcatttgatttagactattttatttctttttgaatattattaaacatgtgttttttctcaatttctgatgtaatgcttttgaaatgctatatgcaatttattgtttgcatttctcaacctcatctgtatcagcttggttgtaaatattaccttcttgtcaatcatattgaccaatcaaacgatgttttgcagtcagatcttaatttttttaatgtttcaattggctgataaacattatagcagtgtaaataagacgatgttacaacttctcagtcagcttttgagttttcaaaaagactttgattttgttgcactgtttattgttctctgtgagtgtttccgggaggatattgagagacctgcggataatattttgtgtgcgtggtgtggcacatccggcctccctctttcatcccagctgcagtcaggttgggagacagggagcattcagcctgaacactggggcttgtgtgctgaacgctggacctctctggaatgtgtttgtgcgtgagagagagatggagttgagagagttggcgctgtttgattaaccatgttgctctgtgtgtgtgtgtttcagattagtaagctgtgcccggctgtgagaaaatcatgtcaagcaagggcgaggagGACAAtATGGCCTCtCTGCCCGCAACACACGAGCTGCATATTTTCGGAAGCATCAAcGGcGTGGATTTCGAtATGGTtGGgCAaGGaACTGGAAACCCAAATGAcGGaTACGAGGAACTGAATCTGAAGTCAACCAAAGGCGACCTCCAaTTCtcaCCTTGGATTCTcGTtCCCCAtATTGGCTATGGaTTTCATCAaTATCTGCCaTAtCCTGAtGGAATGTCACCATTTCAaGCcGCTATGGTGGATGGATCTGGCTACCAaGTCCACcgcACCATGCAaTTTGAGGACGGcGCCtccCTGACTGTGAACTACCGCTATACCTACGAGGGATCtCATATCAAGGGCGAAGCACAaGTtAAAGGaACAGGATTCCCAGCTGAcGGCCCCGTCATGACAAACTCTCTGACCGCCGCCGACTGGAGCCGGTCCAAGAAAACTTACCCTAACGATAAGACCATCATCTCTACCTTCAAaTGGAGTTATACCACcGGCAACGGaAAGCGcTACAGAAGCACAGCCCGAACTACCTATACTTTTGCTAAGCCtATGGCTGCAAACTATCTGAAAAATCAGCCTATGTAtGTcTTCcGAAAaACcGAattgAAGCACTCCAAAACAGAACTGAATTTCAAGGAgTGGCAGAAGGCTTTTACCGATGTtATGGGcatggacgagctgtacaaaggcGGatccggtggatccggtggatctgctacccagtgtaagttccctttacatttttttttcaccgttttaaaaacttaaattggttaaagtgacgggttatccatttttttatgtgcagctgacttgatggagctggaaatggccatggatccggatcgcaaggctgcagtcagccattggcagcaacagtcttacctggactcaggaatccactctggggccacaactactgccccgtccctgagtggcaaaggcaacccggaggatgacgatgtggataatcaggtgctttatgagtgggagcagggcttcaaccagtccttcaaccaagagcaagtagcaggtgtgtggccattctgttgctttagtgaatatctgagaattaaatgtgtaggaaaataagtattctgtttacttaaagggatggttcattcaaaactgaaaattctgccattatttagtcatcttttacttgccacaaacctgtttgagtttctttttttttgttaaacacaaaagaagtagctttggatgatgttggaaaccagaaaccattgacatccgtaatagaacaaaaaaaaaagatactacggaagtcagtgattacaggtttctaacaatgttgaaatcttttgtgttcaacagaaaaataaaaccatcaaatgtatcatag

Table 3.2: plasmid backbone

Feature	**Sequence (5’ to 3’, descending)**

biotinylated-oligonucleotide amplification end side	CACGGTCTCGGCATTCCTGCTGAAC
remainder of backbone	ctggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttaccc aacttaatcgccttgcagcacatccccctttcgccagctggcgtaatag cgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctg aatggcgaatgggacgcgccctagcgcccgctcctttcgctttcttccct tcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggg gctcccttagggttccgatttagtgctttacggcacctcgaccccaaaa aacttgattagggtgatggttcacgtagtgggccatcgccctgatagac ggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgtt ccaaactggaacaacactcaaccctatctcggtctattcttttgatttata agggattttgccgatttcggcctattggttaaaaaatgagctgatttaaca aaaatttaacgcgaattttaacaaaatattaacgcttacaatttaggtgg cacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatac attcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaa taatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgccct tattccctttttttgcggcattttgccttcctgtttttgctcacccagaaacgctg gtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttac atcgaactggatctcaacagcggtaagatccttgagagttttcgccccg aagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcg gtattatcccgtattgacgccgggcaagagcaactcggtcgccgcata cactattctcagaatgacttggttgagtactcaccagtcacagaaaagc atcttacggatggcatgacagtaagagaattatgcagtgctgccataa ccatgagtgataacactgcggccaacttacttctgacaacgatcggag gaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaa ctcgccttgatcgttgggaaccggagctgaatgaagccataccaaac gacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgc gcaaactattaactggcgaactacttactctagcttcccggcaacaatt aatagactggatggaggcggataaagttgcaggaccacttctgcgctc ggcccttccggctggcttcagtgagcgaggaagcggaagagcgccc aatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcag ctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaac gcaattaatgtgagttagctcactcattaggcaccccaggctttacacttt atgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcac acaggaaacagctatgaccatgattacgccaagcgcgcaattaacc ctcactaaag
biotinylated-oligonucleotide amplification start side	GCATACGAGATCTCTTTCCCTACACG

Table 3.3: sgRNAs for genome editing

gene	tag terminus	sgRNA (IDT) + PAM
ap2s1	C	Bottom strand: CTGATTGACAGTTTACTCCAGGG
arpc3	C	Bottom strand: TGAGTAGGTTCTTACTGGCCCGG
cav1	C	Top strand ccgggtcactgctactaaggtTg
clta	C	Bottom strand: GAAGCTGTTTAGCGGACGAGCGG
ctnnb1	N	Top strand: TCAGATTAGTAAGCTGTGCCCGG

Fluorescence Imaging

Dechorionated embryos were mounted in 1.2% agarose canyons and maintained in Egg Water (6 g Instant Ocean, 1.5 g CaS04, 1 L of water) and imaged at 28.5°C. Embryos were immobilized with ~1 ng of α-bungarotoxin protein (TOCHRIS) injected into the common cardinal vein or by submerging embryos in 1-3 mg/mL of tricaine (0.25 g of Tricaine powder, 244.5 ml of Egg Water (500 ml water, 2.5 ml of 1 M Tris, pH 8.0, 0.25 ml of 1 M NaOH, pH to 7.0). ~1 ng of each labeled transferrin (tfa-AlexaFluor647/Jackson ImmunoResearch, tfa-AlexaFluor488/Jackson ImmunoResearch, tfa-pHrodo/Thermo Fisher) was co-injected with α-bungarotoxin. All imaging was performed on an Zeiss 980 laser scanning confocal microscopy equipped Airyscan with either a W Plan-apochromat 20x/1,0 DIC (UV)Vis-IR M27 75 mm objective and LD LCI Plan-Apochromat 40x/1.2 Imm Korr DIC M27 objective. All imaging conditions (laser, dwell, scanning direction, digital gain, and objective) can be found in Table 3.4.

Table 3.4: imaging conditions for data in Chapter 3

figure	tags	lifestage (hpf)	lasers (nm for	dwell (micros	scanning	averaging	Airyscan or	digital gain	objective

			wavelengths, % for strength)	econds)	directional		LSM		
3.4 A	cav1-mNG F1 (F0 ox Capser)	24	514, 4	1	bidirectional	1	Airyscan	910	40x
3.4 B	cav1-mNG F1 (F0 ox Capser)	24	514, 3	2	bidirectional	1	Airyscan	910	40x
3.5	mNG-ctnnb1 F1 (F0 ox Casper)	13	488, 5	16	unidirectional	1	Airyscan	930	40x
3.6 A-D	mNG-ctnnb1 homoix	50	488, 1	2	bidirectional	2	Airyscan	920	40x
3.7 A	mNG-ctnnb1 homoix	16	514, 6	34	bidirectional	1	Airyscan	930	40x
3.7 B	mNG-ctnnb1 homoix	16	514, 6	34	bidirectional	1	Airyscan	930	40x
3.7 C	mNG-ctnnb1 homoix	24	488, 0.2	6	bidirectional	2	LSM	650	40x

3.7 D	mNG-ctnnb1 homoix	24	514, 2	2	bidirectional	1	LSM	880	40x
3.8 A	arpc3-mNG homoix	30	514, 2	3	bidirectional	2	Airyscan	930	40x
3.8 B	arpc3-mNG homoix	50	514, 2	3	bidirectional	2	Airyscan	930	40x
3.8 C	arpc3-mNG homoix	30	488, 3	3	bidirectional	2	Airyscan	930	40x
3.9 A	arpc3-mNG homoix	50	514, 3	17	bidirectional	1	Airyscan	930	40x
3.9 B	arpc3-mNG homoix	55	514,1.2	1	bidirectional	2	Airyscan	930	40x
3.10 A	arpc3-mNG homoix	24	514, 2	3	bidirectional	2	Airyscan	930	40x
3.10 B	arpc3-mNG homoix	50	514, 3	17	bidirectional	1	Airyscan	930	40x
3.11 A-C	ap2s1-mSI (het) oxarpc3-mNG	72	561, 2, 488, 2	24	bidirectional	1	Airyscan	930	40x

	(het)								
3.12 A	arpc3-mNG homo ix	22	639, 2, 488, 6	16	bidirectional	1	Airyscan	930, 930	40x
3.12 B	arpc3-mNG homo ix	30	488,3	16	bidirectional	2	Airyscan	930	40x
3.12 C	arpc3-mNG homo ix	30	514, 2	5	bidirectional	2	Airyscan	930	40x
3.13	ap2s1-mNG homo (tag and Casper) ix	50 hpf	488, 4.5	9	bidirectional	1	Airyscan	930	40x
3.14	ap2s1-mNG homo (tag and Casper) ix	100	514, 3	1	bidirectional	1	Airscan	930	20x
3.16 A-D	ap2s1-mNG	22	639, 0.2, 488, 4	9	bidirectional	1	Airyscan	850, 930	20x
3.17 A	Casper ix	28	561, 2, 488, 2	1	bidirectional	2	Airyscan	930, 930	20x
3.17 B	Casper ix	55	561, 2, 488, 2	1	bidirectional	2	Airyscan	930, 930	20x
3.18 A	ap2s1-mSI F1 female	30 hpf	561, 3, 488, 3	3	bidirectional	1	Airyscan	930,930	40x

	(het) ox clta-mN G F1 male (het)								
3.18 B	ap2s1-mSI F1 female (het) ox clta-mN G F1 male (het)	72	561, 3, 488, 3	48	bidirecti onal	1	LSM	930, 900	40x
3.19	ap2s1-mNG homo (tag and Casper) ix	30	639, 2, 488, 4	21	bidirecti onal	1	Airysca n	930, 930	40x
3.20	ap2s1-mSI (het), clta-mN G (het), Casper, ix	80	561, 1.5, 488, 2.5	2	bidirecti onal	1	Airysca n	930, 930	20x

References

1. Mettlen, M., Chen, P.-H., Srinivasan, S., Danuser, G. & Schmid, S. L. Regulation of Clathrin-Mediated Endocytosis. *Annu. Rev. Biochem.* **87**, 871–896 (2018).
2. Tagliatti, E. & Cortese, K. Imaging Endocytosis Dynamics in Health and Disease. *Membranes* **12**, 393 (2022).
3. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biol.* **16**, e2005970 (2018).

4. Viana, M. P. *et al.* Integrated intracellular organization and its variations in human iPS cells. *Nature* **613**, 345–354 (2023).
5. Bagheri, N., Carpenter, A. E., Lundberg, E., Plant, A. L. & Horwitz, R. The new era of quantitative cell imaging—challenges and opportunities. *Mol. Cell* **82**, 241–247 (2022).
6. Li, D. *et al.* Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. *Science* **349**, aab3500 (2015).
7. Doyon, J. B. *et al.* Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nat. Cell Biol.* **13**, 331–337 (2011).
8. Sun, Y. *et al.* Direct comparison of clathrin-mediated endocytosis in budding and fission yeast reveals conserved and evolvable features. *eLife* **8**, e50749 (2019).
9. He, K. *et al.* Dynamics of Auxilin 1 and GAK in clathrin-mediated traffic. *J. Cell Biol.* **219**, e201908142 (2020).
10. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
11. Roberts, B. *et al.* Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. *Mol. Biol. Cell* **28**, 2854–2874 (2017).
12. Kaksonen, M., Sun, Y. & Drubin, D. G. A Pathway for Association of Receptors, Adaptors, and Actin during Endocytic Internalization. *Cell* **115**, 475–487 (2003).
13. Aguet, F., Antonescu, C. N., Mettlen, M., Schmid, S. L. & Danuser, G. Advances in Analysis of Low Signal-to-Noise Images Link Dynamin and AP2 to the Functions of an Endocytic Checkpoint. *Dev. Cell* **26**, 279–291 (2013).
14. Loerke, D. *et al.* Cargo and Dynamin Regulate Clathrin-Coated Pit Maturation. *PLOS Biol.* **7**, e1000057 (2009).
15. Jin, M. *et al.* Branched actin networks are organized for asymmetric force production during clathrin-mediated endocytosis in mammalian cells. *Nat. Commun.* **13**, 3578 (2022).
16. Taylor, M. J., Perrais, D. & Merrifield, C. J. A High Precision Survey of the Molecular

Dynamics of Mammalian Clathrin-Mediated Endocytosis. *PLOS Biol.* **9**, e1000604 (2011).

17. Drubin, D. G. & Hyman, A. A. Stem cells: the new “model organism”. *Mol. Biol. Cell* **28**, 1409–1411 (2017).

18. Hong, S. H., Cortesio, C. L. & Drubin, D. G. Machine-Learning-Based Analysis in Genome-Edited Cells Reveals the Efficiency of Clathrin-Mediated Endocytosis. *Cell Rep.* **12**, 2121–2130 (2015).

19. Roux, A. *et al.* Membrane curvature controls dynamin polymerization. *Proc. Natl. Acad. Sci.* **107**, 4141–4146 (2010).

20. Rottner, K., Faix, J., Bogdan, S., Linder, S. & Kerkhoff, E. Actin assembly mechanisms at a glance. *J. Cell Sci.* **130**, 3427–3435 (2017).

21. Lu, R., Drubin, D. G. & Sun, Y. Clathrin-mediated endocytosis in budding yeast at a glance. *J. Cell Sci.* **129**, 1531–1536 (2016).

22. Aghamohammadzadeh, S. & Ayscough, K. R. Differential requirements for actin during yeast and mammalian endocytosis. *Nat. Cell Biol.* **11**, 1039–1042 (2009).

23. Lacy, M. M., Ma, R., Ravindra, N. G. & Berro, J. Molecular mechanisms of force production in clathrin-mediated endocytosis. *FEBS Lett.* **592**, 3586–3605 (2018).

24. Hassinger, J. E., Oster, G., Drubin, D. G. & Rangamani, P. Design principles for robust vesiculation in clathrin-mediated endocytosis. *Proc. Natl. Acad. Sci.* **114**, E1118–E1127 (2017).

25. Lanzetti, L. Actin in membrane trafficking. *Curr. Opin. Cell Biol.* **19**, 453–458 (2007).

26. Boulant, S., Kural, C., Zeeh, J.-C., Ubelmann, F. & Kirchhausen, T. Actin dynamics counteract membrane tension during clathrin-mediated endocytosis. *Nat. Cell Biol.* **13**, 1124–1131 (2011).

27. Kaplan, C. *et al.* Load adaptation by endocytic actin networks. *Mol. Biol. Cell* **33**, ar50 (2022).

28. Batchelder, E. M. & Yarar, D. Differential Requirements for Clathrin-dependent

Endocytosis at Sites of Cell–Substrate Adhesion. *Mol. Biol. Cell* **21**, 3070–3079 (2010).

29. Kaur, S., Fielding, A. B., Gassner, G., Carter, N. J. & Royle, S. J. An unmet actin requirement explains the mitotic inhibition of clathrin-mediated endocytosis. *eLife* **3**, e00829 (2014).

30. Mechanoregulation of clathrin-mediated endocytosis | Journal of Cell Science | The Company of Biologists. https://journals.biologists.com/jcs/article/130/21/3631/56414/Mechanoregulation-of-clathrin-mediated-endocytosis.

31. Grassart, A. *et al.* Actin and dynamin2 dynamics and interplay during clathrin-mediated endocytosis. *J. Cell Biol.* **205**, 721–735 (2014).

32. Akamatsu, M. *et al.* Principles of self-organization and load adaptation by the actin cytoskeleton during clathrin-mediated endocytosis. *eLife* **9**, e49840 (2020).

33. Mund, M. *et al.* Systematic Nanoscale Analysis of Endocytosis Links Efficient Vesicle Formation to Patterned Actin Nucleation. *Cell* **174**, 884-896.e17 (2018).

34. Yarar, D., Waterman-Storer, C. M. & Schmid, S. L. A Dynamic Actin Cytoskeleton Functions at Multiple Stages of Clathrin-mediated Endocytosis. *Mol. Biol. Cell* **16**, 964–975 (2005).

35. Collins, A., Warrington, A., Taylor, K. A. & Svitkina, T. Structural Organization of the Actin Cytoskeleton at Sites of Clathrin-Mediated Endocytosis. *Curr. Biol.* **21**, 1167–1175 (2011).

36. Cureton, D. K., Massol, R. H., Saffarian, S., Kirchhausen, T. L. & Whelan, S. P. J. Vesicular Stomatitis Virus Enters Cells through Vesicles Incompletely Coated with Clathrin That Depend upon Actin for Internalization. *PLOS Pathog.* **5**, e1000394 (2009).

37. Bucher, D. *et al.* Clathrin-adaptor ratio and membrane tension regulate the flat-to-curved transition of the clathrin coat during endocytosis. *Nat. Commun.* **9**, 1109 (2018).

38. Saleem, M. *et al.* A balance between membrane elasticity and polymerization energy sets the shape of spherical clathrin coats. *Nat. Commun.* **6**, 6249 (2015).

39. Liu, A. P., Loerke, D., Schmid, S. L. & Danuser, G. Global and Local Regulation of Clathrin-Coated Pit Dynamics Detected on Patterned Substrates. *Biophys. J.* **97**, 1038–1047 (2009).

40. Yoshida, A. *et al.* Morphological changes of plasma membrane and protein assembly during clathrin-mediated endocytosis. *PLOS Biol.* **16**, e2004786 (2018).

41. Cheng, X. *et al.* Dynamin-dependent vesicle twist at the final stage of clathrin-mediated endocytosis. *Nat. Cell Biol.* **23**, 859–869 (2021).

42. Dambournet, D. *et al.* Genome-edited human stem cells expressing fluorescently labeled endocytic markers allow quantitative analysis of clathrin-mediated endocytosis during differentiation. *J. Cell Biol.* **217**, 3301–3311 (2018).

43. Ferguson, S. *et al.* Coordinated Actions of Actin and BAR Proteins Upstream of Dynamin at Endocytic Clathrin-Coated Pits. *Dev. Cell* **17**, 811–822 (2009).

44. Durrbach, A., Louvard, D. & Coudrier, E. Actin filaments facilitate two steps of endocytosis. *J. Cell Sci.* **109**, 457–465 (1996).

45. Lamaze, C., Fujimoto, L. M., Yin, H. L. & Schmid, S. L. The Actin Cytoskeleton Is Required for Receptor-mediated Endocytosis in Mammalian Cells*. *J. Biol. Chem.* **272**, 20332–20335 (1997).

46. Fujimoto, L. M., Roth, R., Heuser, J. E. & Schmid, S. L. Actin Assembly Plays a Variable, but not Obligatory Role in Receptor-Mediated Endocytosis. *Traffic* **1**, 161–171 (2000).

47. Serwas, D. *et al.* Mechanistic insights into actin force generation during vesicle formation from cryo-electron tomography. *Dev. Cell* **57**, 1132-1145.e5 (2022).

48. Xu, R. & Du, S. Overexpression of Lifeact-GFP Disrupts F-Actin Organization in Cardiomyocytes and Impairs Cardiac Function. *Front. Cell Dev. Biol.* **9**, (2021).

49. Flores, L. R., Keeling, M. C., Zhang, X., Sliogeryte, K. & Gavara, N. Lifeact-TagGFP2 alters F-actin organization, cellular morphology and biophysical behaviour. *Sci. Rep.* **9**, 3241 (2019).

50. Grossier, J.-P., Xouri, G., Goud, B. & Schauer, K. Cell adhesion defines the topology of endocytosis and signaling. *EMBO J.* **33**, 35–45 (2014).

51. Grimm, J. B. *et al.* A general method to fine-tune fluorophores for live-cell and in vivo imaging. *Nat. Methods* **14**, 987–994 (2017).

52. 4D cell biology: big data image analytics and lattice light-sheet imaging reveal dynamics of clathrin-mediated endocytosis in stem cell–derived intestinal organoids | Molecular Biology of the Cell. https://www.molbiolcell.org/doi/10.1091/mbc.E18-06-0375.

53. Observing the cell in its native state: Imaging subcellular dynamics in multicellular organisms | Science. https://www.science.org/doi/10.1126/science.aaq1392?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed.

54. Wang, K. *et al.* Rapid adaptive optical recovery of optimal resolution over large volumes. *Nat. Methods* **11**, 625–628 (2014).

55. Membrane dynamics of dividing cells imaged by lattice light-sheet microscopy | Molecular Biology of the Cell. https://www.molbiolcell.org/doi/full/10.1091/mbc.E16-03-0164.

56. Swinburne, I. A., Mosaliganti, K. R., Green, A. A. & Megason, S. G. Improved Long-Term Imaging of Embryos with Genetically Encoded α-Bungarotoxin. *PLOS ONE* **10**, e0134005 (2015).

57. Lamellar projections in the endolymphatic sac act as a relief valve to regulate inner ear pressure | eLife. https://elifesciences.org/articles/37131.

58. Höijer, I. *et al.* CRISPR-Cas9 induces large structural variants at on-target and off-target sites in vivo that segregate across generations. *Nat. Commun.* **13**, 627 (2022).

59. Prill, K. & Dawson, J. F. Homology-Directed Repair in Zebrafish: Witchcraft and Wizardry? *Front. Mol. Biosci.* **7**, (2020).

60. Gibbs, E. M. *et al.* Two Dynamin-2 Genes Are Required for Normal Zebrafish Development. *PLOS ONE* **8**, e55888 (2013).

61. Hagemann, A. I. H. *et al.* In vivo analysis of formation and endocytosis of the Wnt/β-catenin signaling complex in zebrafish embryos. *J. Cell Sci.* **127**, 3970–3982 (2014).

62. Cocucci, E., Aguet, F., Boulant, S. & Kirchhausen, T. The First Five Seconds in the Life of a Clathrin-Coated Pit. *Cell* **150**, 495–507 (2012).

63. Kon, S., Tanabe, K., Watanabe, T., Sabe, H. & Satake, M. Clathrin dependent endocytosis of E-cadherin is regulated by the Arf6GAP isoform SMAP1. *Exp. Cell Res.* **314**, 1415–1428 (2008).

64. Ambrosi, G. *et al.* Allele-specific endogenous tagging and quantitative analysis of β-catenin in colorectal cancer cells. *eLife* **11**, e64498 (2022).

65. de Man, S. M., Zwanenburg, G., van der Wal, T., Hink, M. A. & van Amerongen, R. Quantitative live-cell imaging and computational modeling shed new light on endogenous WNT/CTNNB1 signaling dynamics. *eLife* **10**, e66440 (2021).

66. Clevers, H. & Nusse, R. Wnt/β-catenin signaling and disease. *Cell* **149**, 1192–1205 (2012).

67. Rim, E. Y., Clevers, H. & Nusse, R. The Wnt Pathway: From Signaling Mechanisms to Synthetic Modulators. *Annu. Rev. Biochem.* **91**, null (2022).

68. Matthaeus, C. & Taraska, J. W. Energy and Dynamics of Caveolae Trafficking. *Front. Cell Dev. Biol.* **8**, (2021).

69. Induced nanoscale membrane curvature bypasses the essential endocytic function of clathrin | Journal of Cell Biology | Rockefeller University Press. https://rupress.org/jcb/article/221/7/e202109013/213203/Induced-nanoscale-membrane-curvature-bypasses-the.

70. De Ieso, M. L., Kuhn, M., Bernatchez, P., Elliott, M. H. & Stamer, W. D. A Role of Caveolae in Trabecular Meshwork Mechanosensing and Contractile Tone. *Front. Cell Dev. Biol.* **10**, (2022).

71. Hoshijima, K., Jurynec, M. J. & Grunwald, D. J. Precise Editing of the Zebrafish Genome

Made Simple and Efficient. *Dev. Cell* **36**, 654–667 (2016).

72. Dambournet, D., Hong, S. H., Grassart, A. & Drubin, D. G. Tagging endogenous loci for live-cell fluorescence imaging and molecule counting using ZFNs, TALENs, and Cas9. in *Methods in enzymology* vol. 546 139–160 (Elsevier, 2014).

73. Balik-Meisner, M., Truong, L., Scholl, E. H., Tanguay, R. L. & Reif, D. M. Population genetic diversity in zebrafish lines. *Mamm. Genome* **29**, 90–100 (2018).

74. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).

75. Primer Designer. (2022).

76. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

77. Seleit, A., Aulehla, A. & Paix, A. Endogenous protein tagging in medaka using a simplified CRISPR/Cas9 knock-in approach. *eLife* **10**, e75050 (2021).

78. SwinburneLabUCB/Sequencing_File_Renamer. https://github.com/SwinburneLabUCB/Sequencing_File_Renamer.

79. Zhou, Y. *et al.* Caveolin-1 and cavin1 act synergistically to generate a unique lipid environment in caveolae. *J. Cell Biol.* **220**, e202005138 (2021).

80. ZFIN Publication: Thisse et al., 2004. https://zfin.org/ZDB-PUB-040907-1.

81. Egger, A. N. *et al.* The importance of caveolins and caveolae to dermatology: Lessons from the caves and beyond. *Exp. Dermatol.* **29**, 136–148 (2020).

82. Kizil, C. *et al.* Simplet/Fam53b is required for Wnt signal transduction by regulating β-catenin nuclear localization. *Development* **141**, 3529–3539 (2014).

83. Langenberg, T. & Brand, M. Lineage restriction maintains a stable organizer cell population at the zebrafish midbrain-hindbrain boundary. *Development* **132**, 3209–3216 (2005).

84. Riccomagno, M. M., Takada, S. & Epstein, D. J. Wnt-dependent regulation of inner ear

morphogenesis is balanced by the opposing and supporting roles of Shh. *Genes Dev.* **19**, 1612–1623 (2005).

85. Esho, T. *et al.* Anchoring Cords: A Distinct Suprastructure in the Developing Skin. *J. Invest. Dermatol.* **142**, 2940-2948.e2 (2022).

86. Haseeb, M., Pirzada, R. H., Ain, Q. U. & Choi, S. Wnt Signaling in the Regulation of Immune Cell and Cancer Therapeutics. *Cells* **8**, 1380 (2019).

87. Lu, T. Q., Van Loon, A. P. & Sagasti, A. How to wrinkle a cell: Emerging mechanisms of microridge morphogenesis. *Curr. Opin. Cell Biol.* **76**, 102088 (2022).

88. Dewit, J., Witten, P. E. & Huysseune, A. The mechanism of cartilage subdivision in the reorganization of the zebrafish pectoral fin endoskeleton. *J. Exp. Zoolog. B Mol. Dev. Evol.* **316**, 584–597 (2011).

89. Liu, Y., Bao, H., Mei, Y., Han, X. & Ji, P. Formin Links Erythroid Cytoskeleton to Organelle Clearance through Escrt-III Complex during Reticulocyte Maturation. *Blood* **134**, 151 (2019).

90. Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).

91. Wattrus, S. J. & Zon, L. I. Stem cell safe harbor: the hematopoietic stem cell niche in zebrafish. *Blood Adv.* **2**, 3063–3069 (2018).

92. Blaser, B. W. *et al.* CXCR1 remodels the vascular niche to promote hematopoietic stem and progenitor cell engraftment. *J. Exp. Med.* **214**, 1011–1027 (2017).

93. Rodríguez-Fraticelli, A. E. *et al.* Developmental regulation of apical endocytosis controls epithelial patterning in vertebrate tubular organs. *Nat. Cell Biol.* **17**, 241–250 (2015).

94. Kawabata, H. Transferrin and transferrin receptors update. *Free Radic. Biol. Med.* **133**, 46–54 (2019).

95. Wingert, R. A. *et al.* The chianti zebrafish mutant provides a model for erythroid-specific disruption of transferrin receptor 1. *Development* **131**, 6225–6235 (2004).

96. Carlevaro, M. F. *et al.* Transferrin Promotes Endothelial Cell Migration and Invasion: Implication in Cartilage Neovascularization. *J. Cell Biol.* **136**, 1375–1384 (1997).

97. Liu, J. *et al.* Quantitative analysis of murine terminal erythroid differentiation in vivo: novel method to study normal and disordered erythropoiesis. *Blood* **121**, e43–e49 (2013).

98. Schekman, R. & Singer, S. J. Clustering and endocytosis of membrane receptors can be induced in mature erythrocytes of neonatal but not adult humans. *Proc. Natl. Acad. Sci.* **73**, 4075–4079 (1976).

99. Fernandes, H. P., Cesar, C. L. & Barjas-Castro, M. de L. Electrical properties of the red blood cell membrane and immunohematological investigation. *Rev. Bras. Hematol. E Hemoter.* **33**, 297–301 (2011).

100. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).

101. Alamos, S., Reimer, A., Niyogi, K. K. & Garcia, H. G. Quantitative imaging of RNA polymerase II activity in plants reveals the single-cell basis of tissue-wide transcriptional dynamics. *Nat. Plants* **7**, 1037–1049 (2021).

102. Rai, M. R., Li, C., Ghashghaei, H. T. & Greenbaum, A. Deep learning-based adaptive optics for light sheet fluorescence microscopy. *Biomed. Opt. Express* **14**, 2905–2919 (2023).