Figures and figure supplements

Deciphering anomalous heterogeneous intracellular transport with neural networks

Daniel Han *et al*

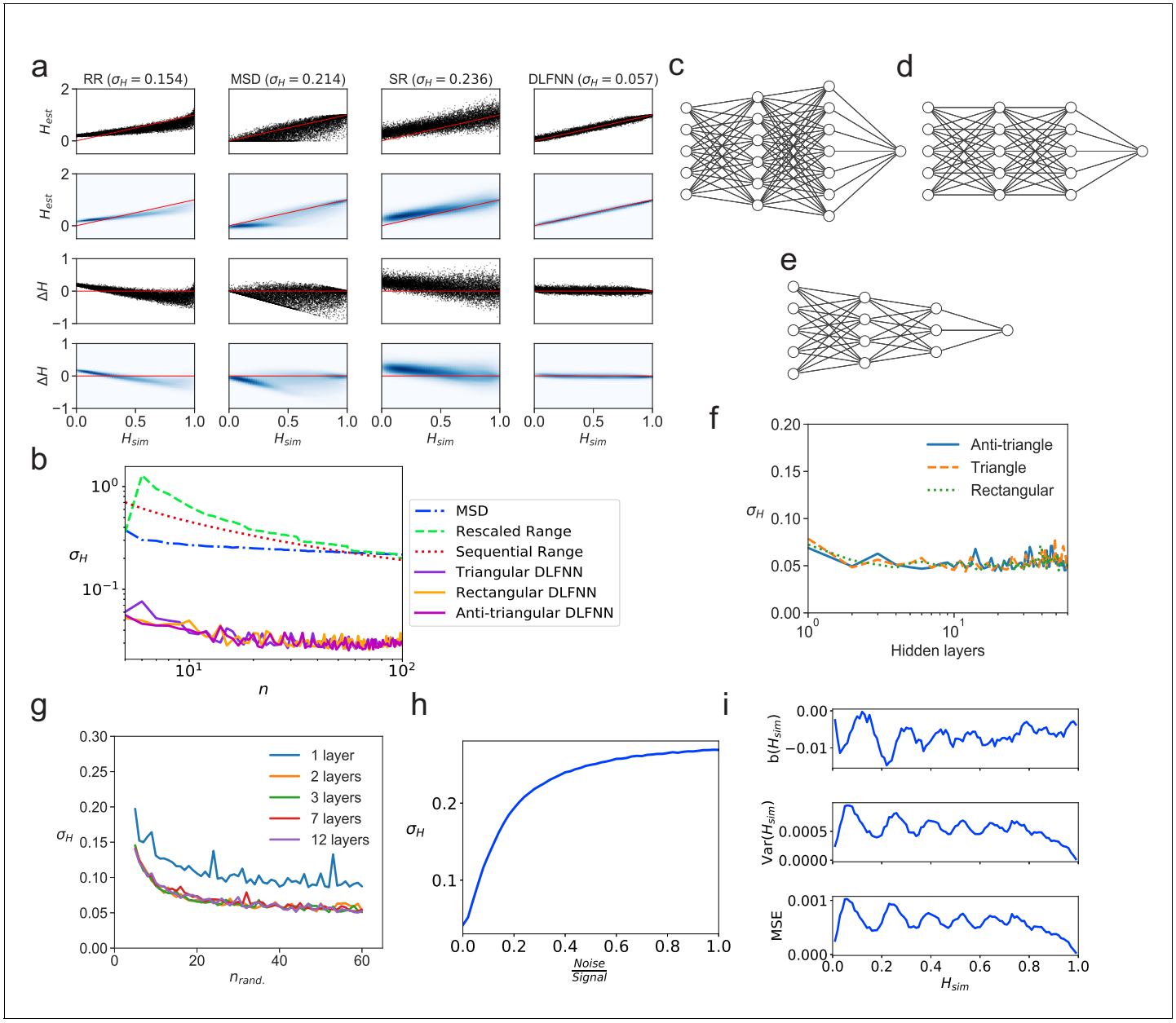


Figure 1. Tests of exponent estimation for the DLFNN using N = 10^4 simulated fBm trajectories. (**a**) Plots showing the Hurst exponent estimates of fBm trajectories with $n = 10^2$ data points by a triangular DLFNN with three hidden layers compared with conventional methods. Plots are vertically grouped by Hurst exponent estimation method: (left to right) rescaled range, MSD, sequential range and DLFNN. σ_H values are shown in the title. *Top row*: Scatter plots of estimated Hurst exponents H_{est} and the true value of Hurst exponents from simulation H_{sim}. The red line shows perfect estimation. *Second row*: Due to the density of points, a Gaussian kernel density estimation was made of the plots in the top row (see Materials and methods). *Third row*: Scatter plots of the difference between the true value of Hurst exponents from simulation and estimated Hurst exponent $\Delta H = H_{sim} - H_{est}$. *Last row*: Gaussian kernel density estimation of the plots in the third row. (**b**) σ_H as a function of the number of consecutive fBm trajectory data points n for different methods of exponent estimation. Example structures for two hidden layers and $n = 5$ time series input points of the anti-triangular, rectangular and triangular DLFNN are shown in (**c, d** and **e**), respectively. (**f**) σ_H as a function of the number of hidden layers in the DLFNN for triangular, rectangular and anti-triangular structures. (**g**) σ_H as a function of the number of randomly sampled fBm trajectory data points n_{rand} with different number of hidden layers in the DLFNN shown in the legend. (**h**) σ_H as a function of the noise-to-signal ratio $(\frac{Noise}{Signal})$ (NSR) from Gaussian random numbers added to all $n = 10^2$ data points in simulated fBm trajectories. (**i**) Plots of bias $b(H_{sim})$, variance $\mathrm{Var}(H_{sim})$ and mean square error (MSE) as functions of H_{sim}. For each value of H_{sim}, fBm trajectories with $n = 100$ points were simulated and estimated by a triangular DLFNN.

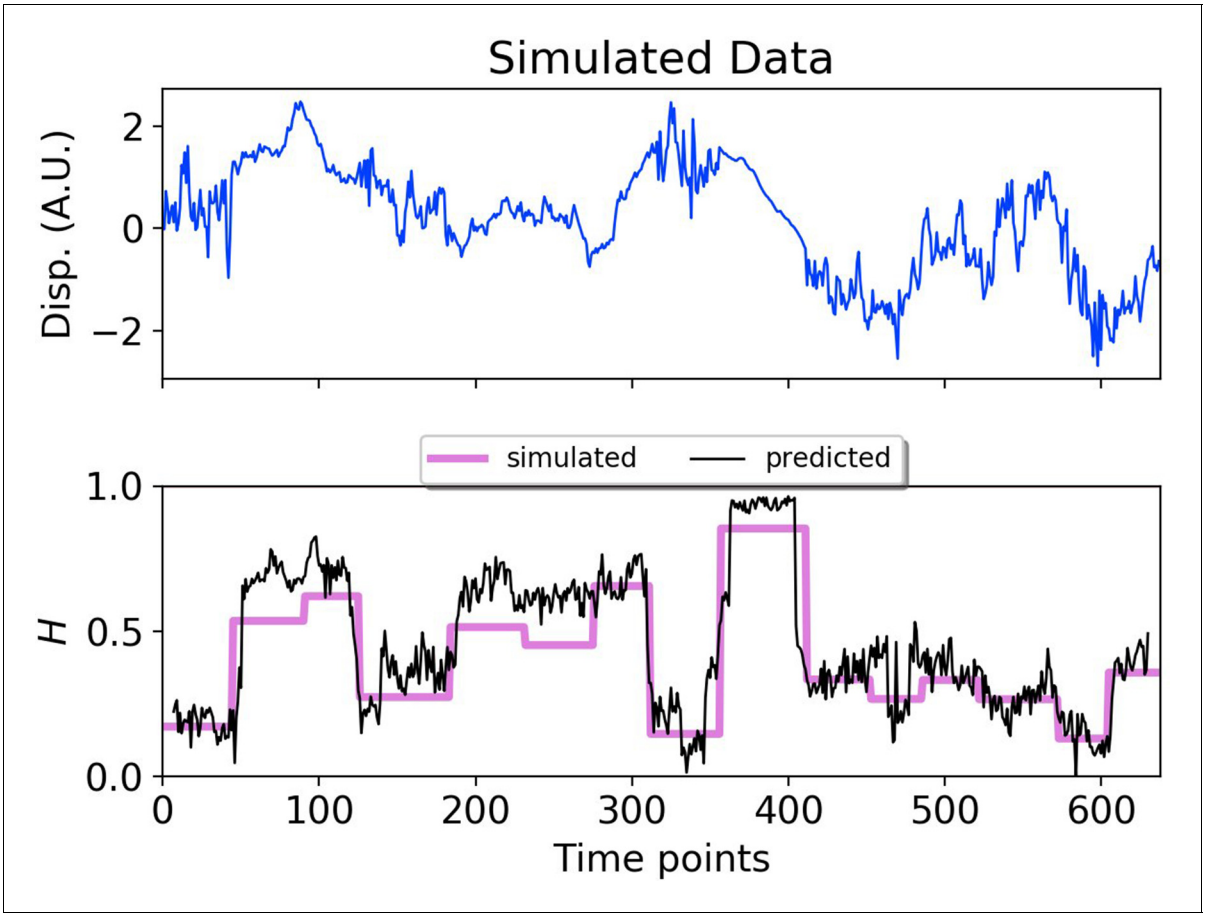


Figure 2. DLFNN analysis of a simulated trajectory. *Top*: Plot of displacement as a function of time from a simulated fBm trajectory (blue) with multiple exponent values. *Bottom*: Hurst exponent values used for simulation (magenta), and the DLFNN exponent predictions of the neural network using a 15 point moving window (black).

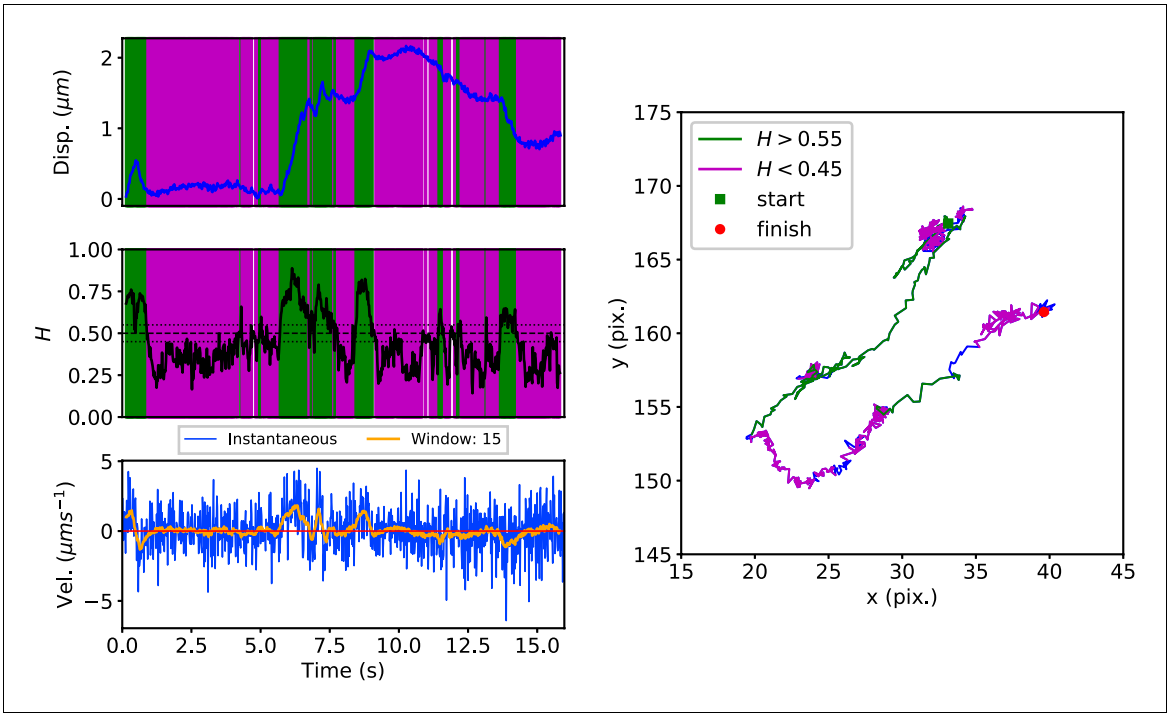


Figure 3. DLFNN analysis of a GFP-Rab5 endosome trajectory. *Top*: Plot of displacement from a single trajectory in an MRC-5 cell (blue). Shaded areas show persistent (0.55 < H < 1 in green) and anti-persistent (0 < H < 0.45 in magenta) behaviour. *Middle*: A 15 point moving window DLFNN exponent estimate for the trajectory (black) with a line (dashed) marking diffusion H = 0.5 and two lines (dotted) marking confidence bounds for estimation marking H = 0.45 and 0.55. *Bottom*: Plot of instantaneous and moving (15 point) window velocity. *Right*: Plot of the trajectory with start and finish positions. Persistent (green) and anti-persistent (magenta) segments are shown. For sections that were 0.45 < H < 0.55 were not classified as persistent or anti-persistent and are depicted in blue.

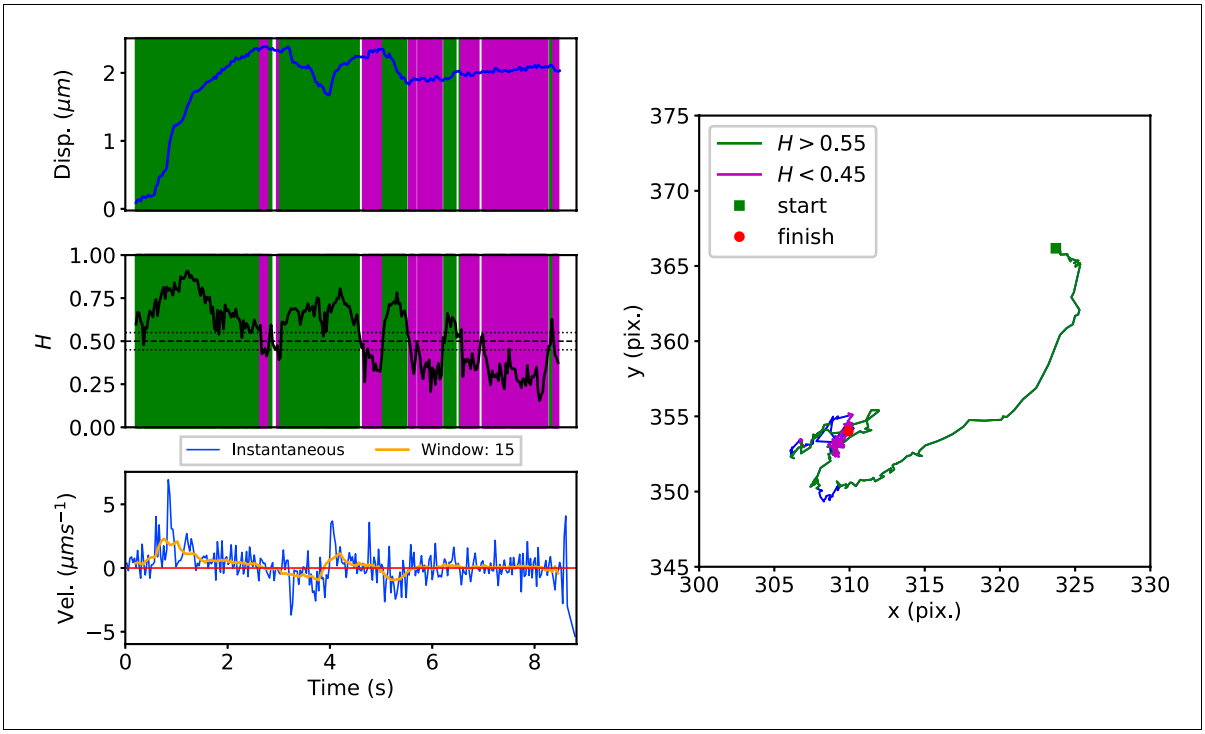


Figure 3—figure supplement 1. DLFNN analysis of a GFP-SNX1-labeled endosome trajectory, depicted as in ***Figure 3***.

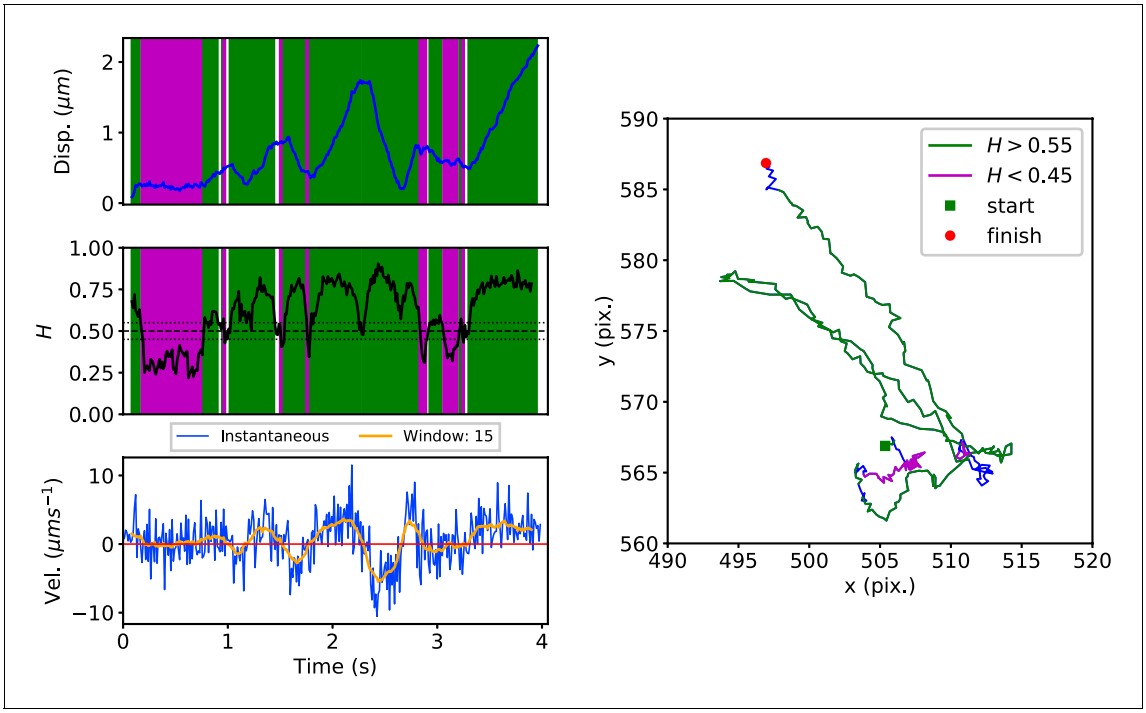


Figure 3—figure supplement 2. DLFNN analysis of a lysosome trajectory, depicted as in ***Figure 3***.

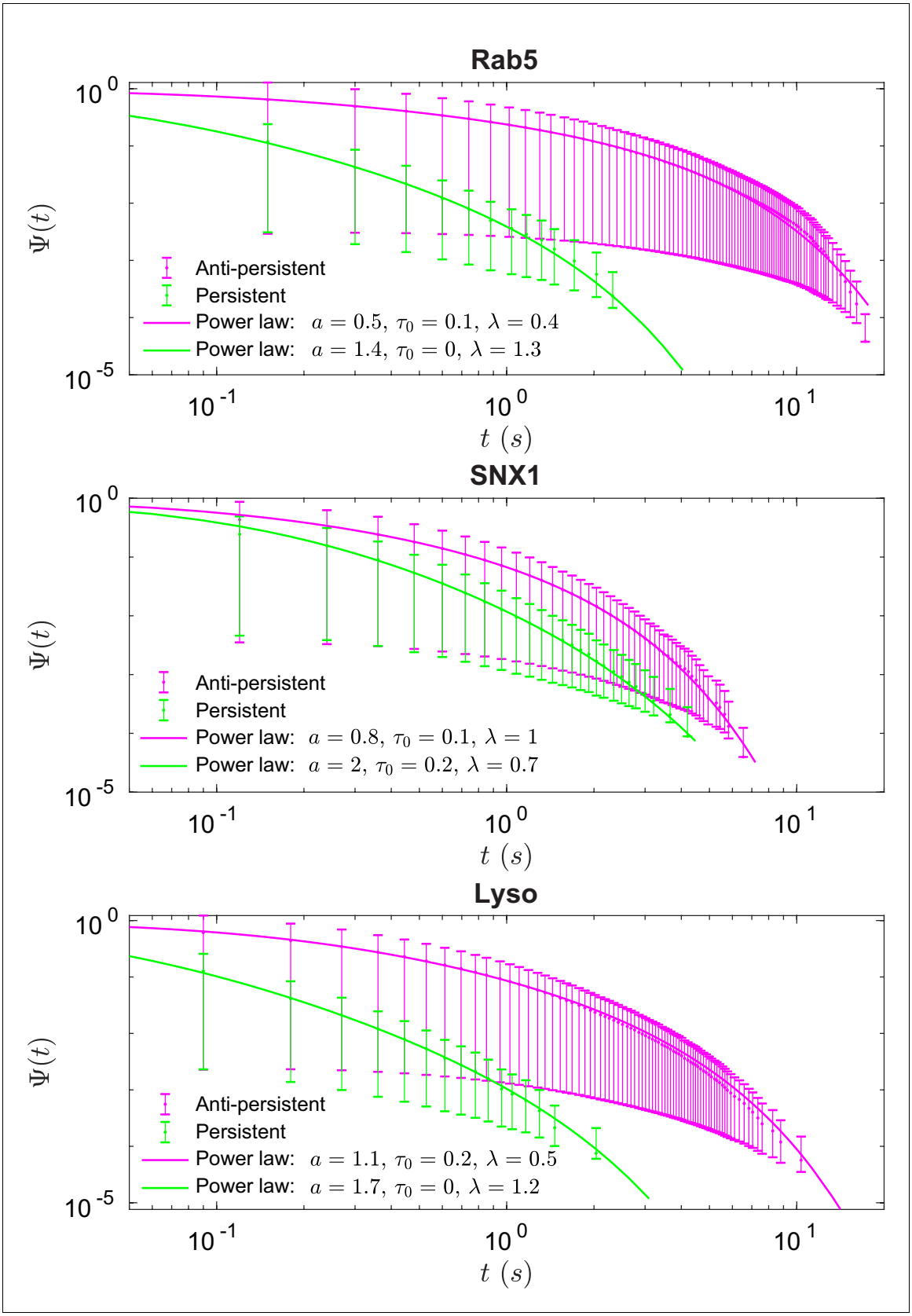


Figure 4. Survival functions plotted with error bars for persistent and anti-persistent segments for Rab5-positive endosomes, SNX1-positive endosomes and lysosomes with the power-law fits. Fit parameters can be found in ***Appendix 3—table 1***.

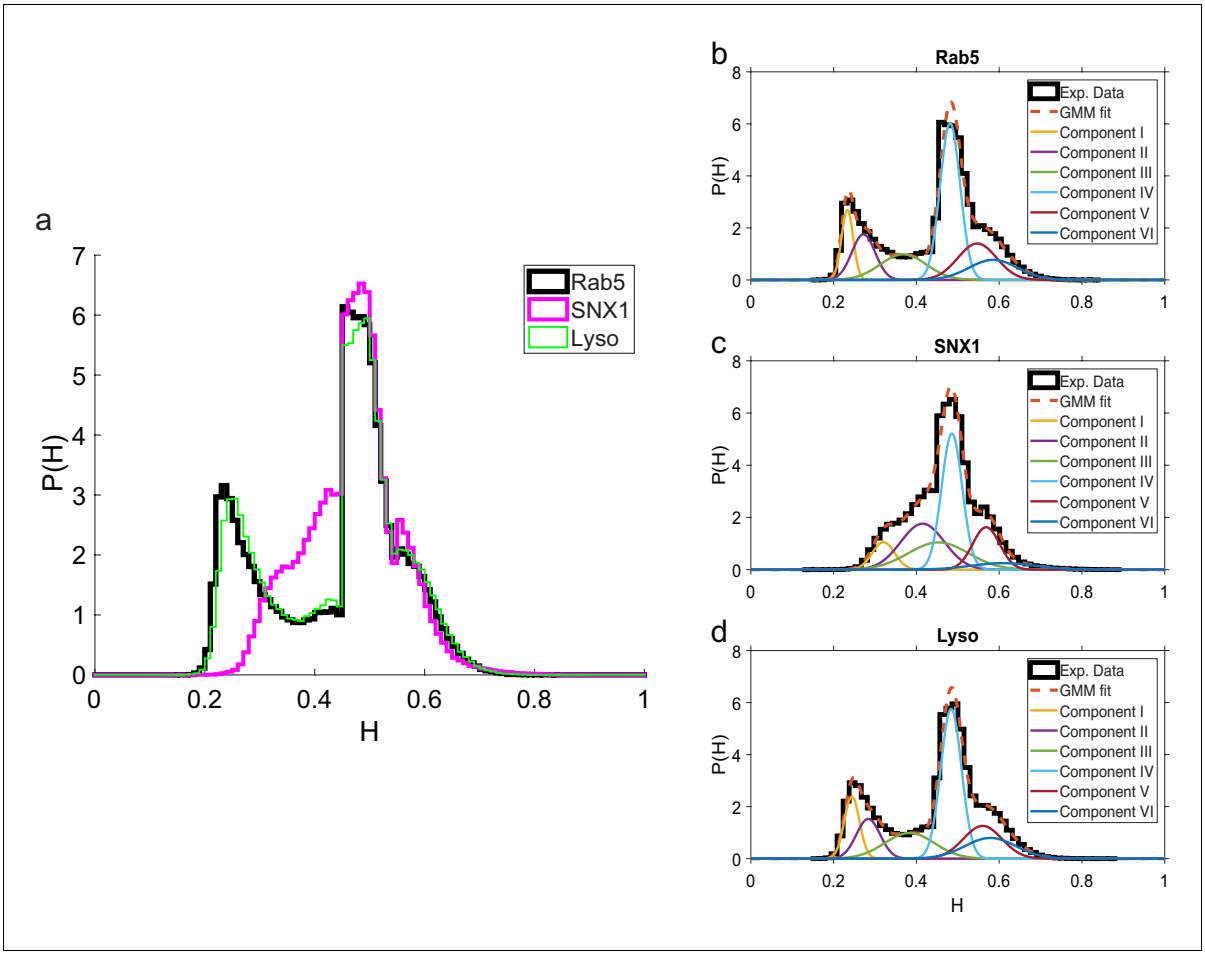


Figure 5. Comparison of Hurst exponent distributions for GFP-Rab5, GFP-SNX1 and lysosomes. (**a**) Histograms of Hurst exponents for GFP-Rab5 (black), GFP-SNX1 (magenta) endosomes and lysosomes (green) plot on the same axes for comparison. The individual histograms of Hurst exponents (black solid) for GFP-Rab5-tagged endosomes, GFP-SNX1-tagged endosomes and lysosomes are shown in (**b, c** and **d**) respectively. For each histogram, the Gaussian mixture model fit for six components (red dashed) and individual Gaussian distribution components are shown on the same plot. The number of components were chosen through the Bayes information criterion shown in ***Appendix 4—figure 1***.

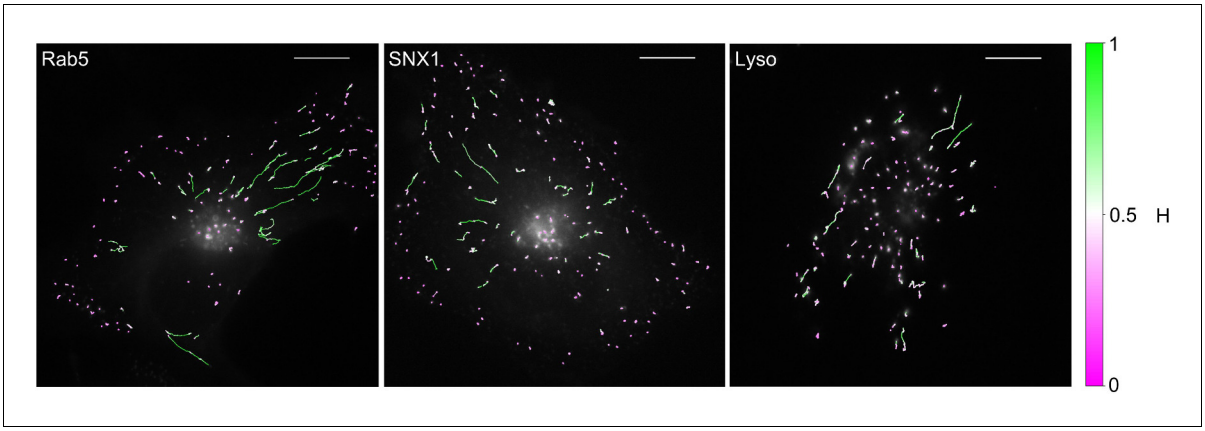


Figure 6. MRC-5 cells stably expressing GFP-Rab5, GFP-SNX1 or stained with Lysobrite with tracking data overlaid. The colours show the value of H estimated by the neural network using a 15 point window. The scalebar is 10 μm.

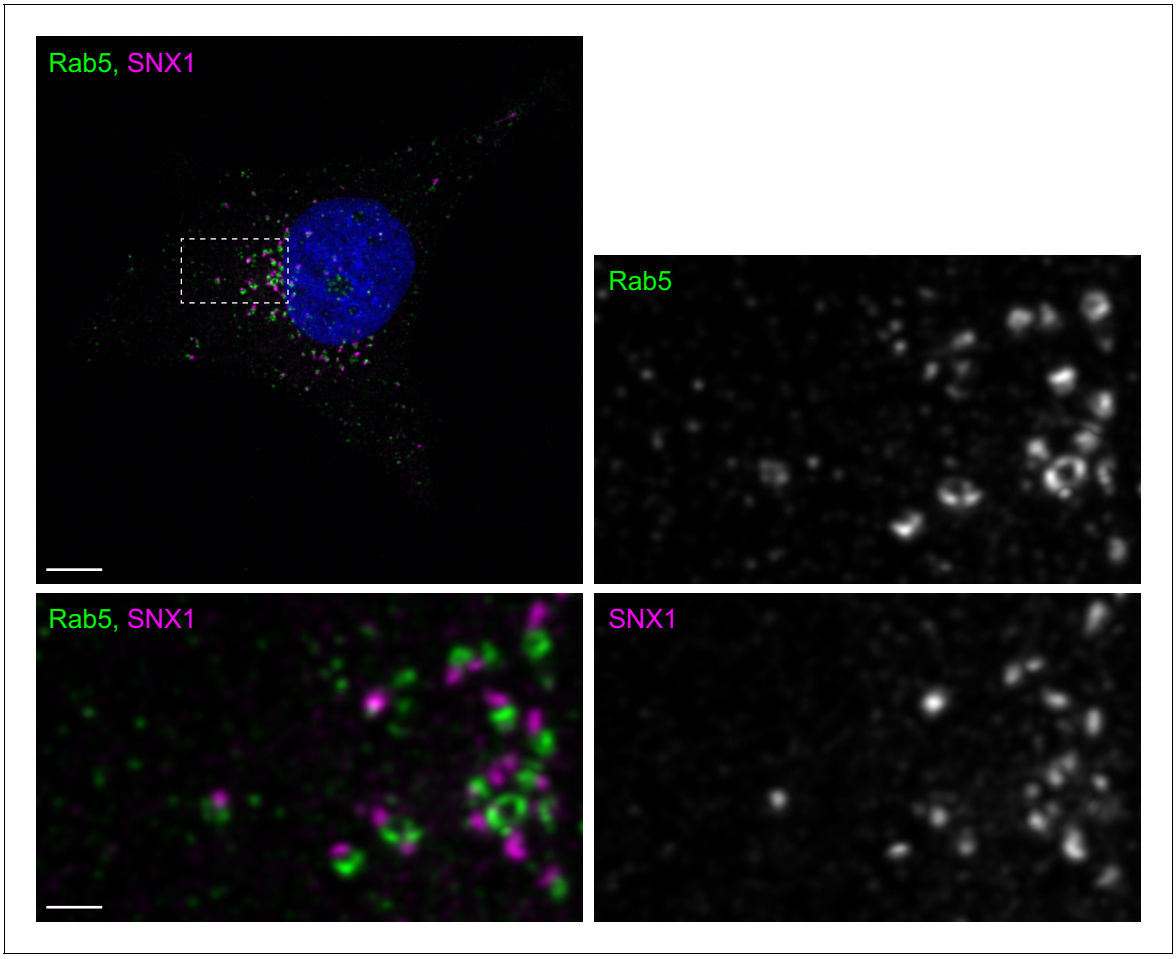


Figure 6—figure supplement 1. Distribution of endogenous Rab5 and SNX1. MRC-5 cells were fixed and labeled with antibodies to Rab5 and SNX1, then imaged by confocal microscopy. A maximum-intensity z-projection of deconvolved images is shown. The boxed region is enlarged and presented as grey-scale single channels and a two color merged image. The scale bar is 10 μm (main image) and 2 μm (enlargements).

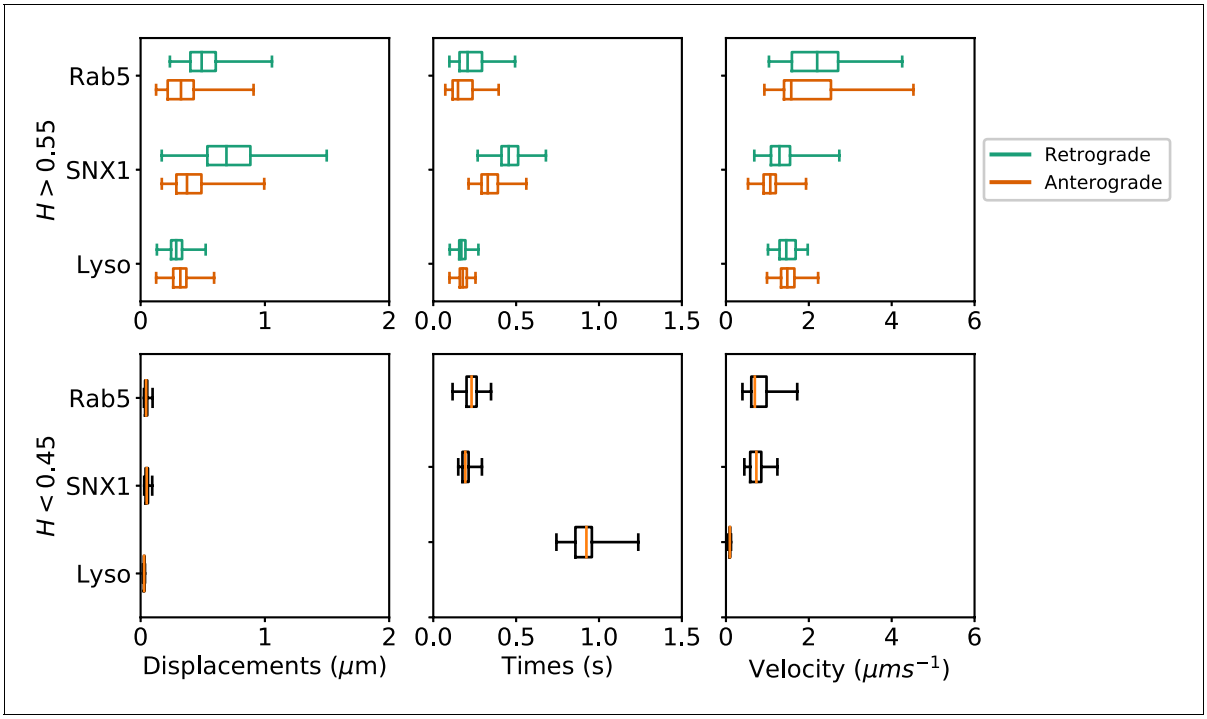
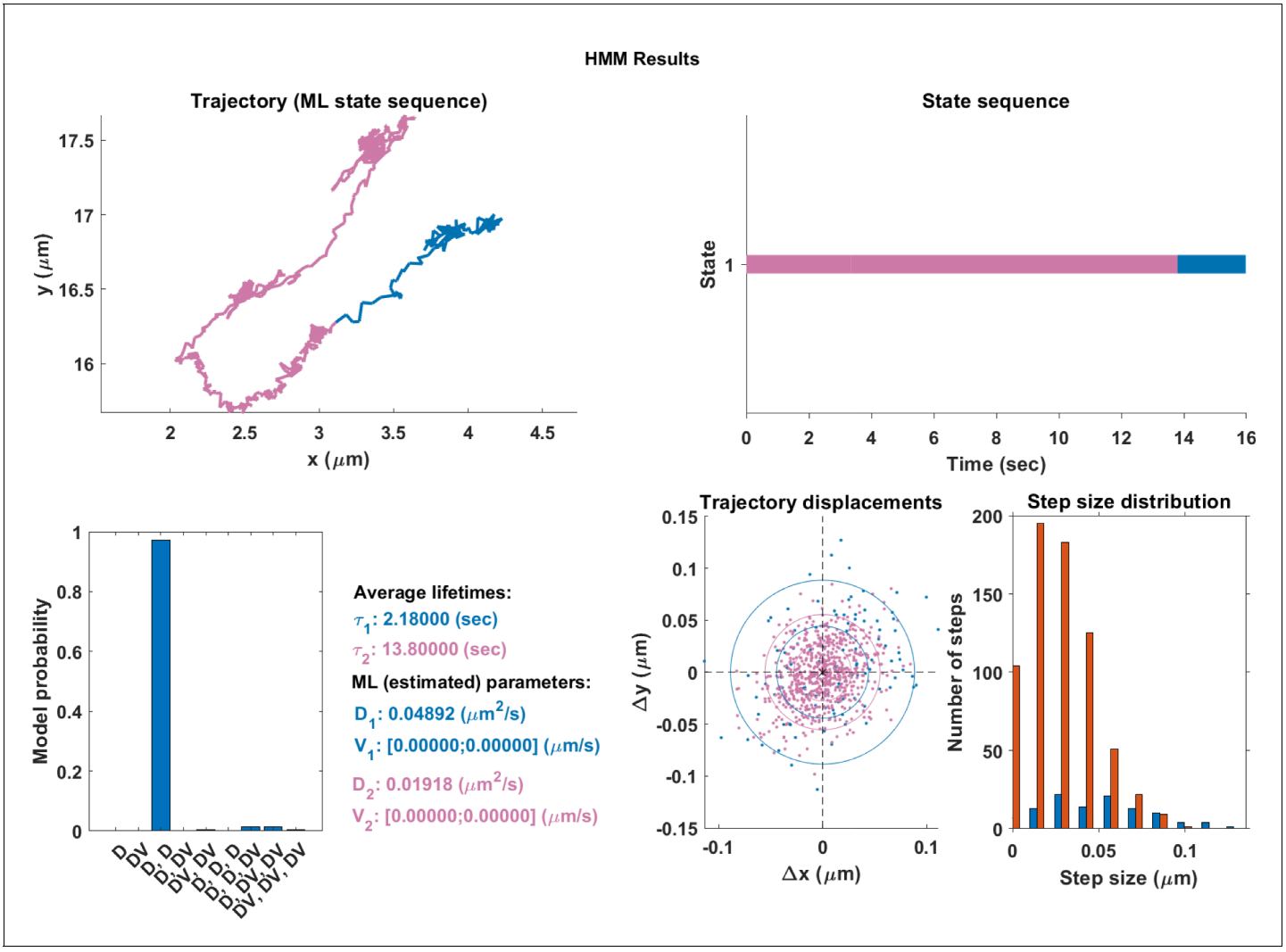
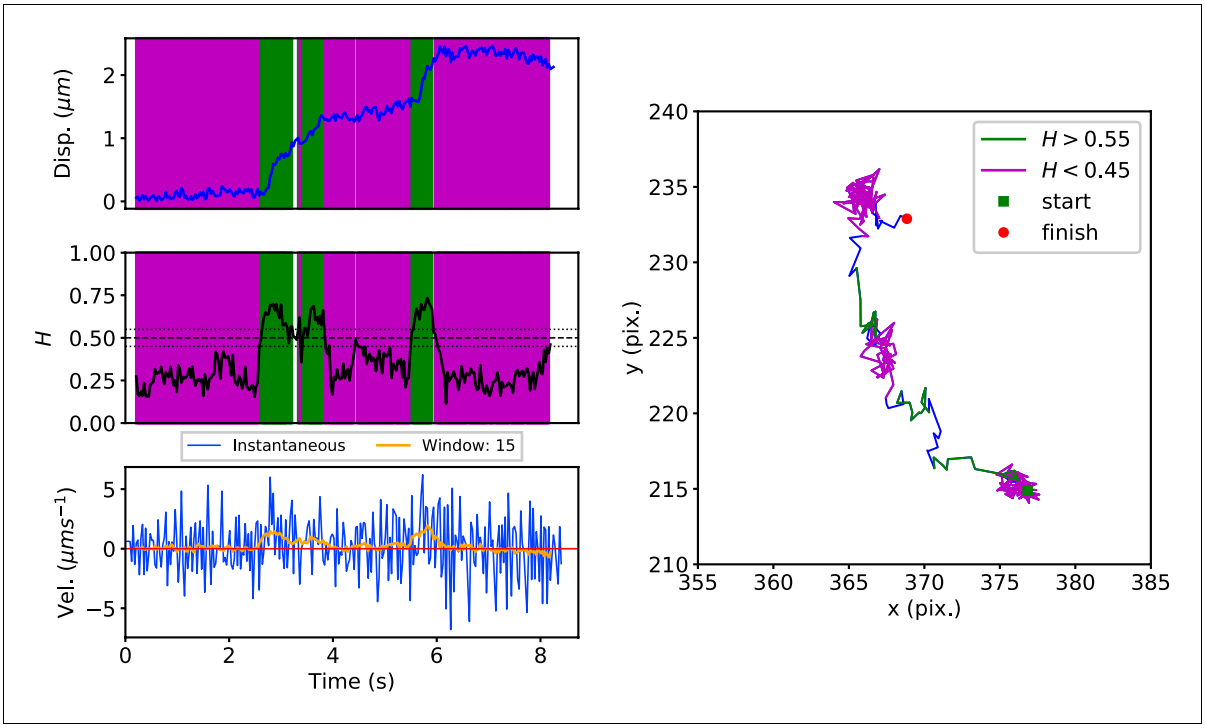


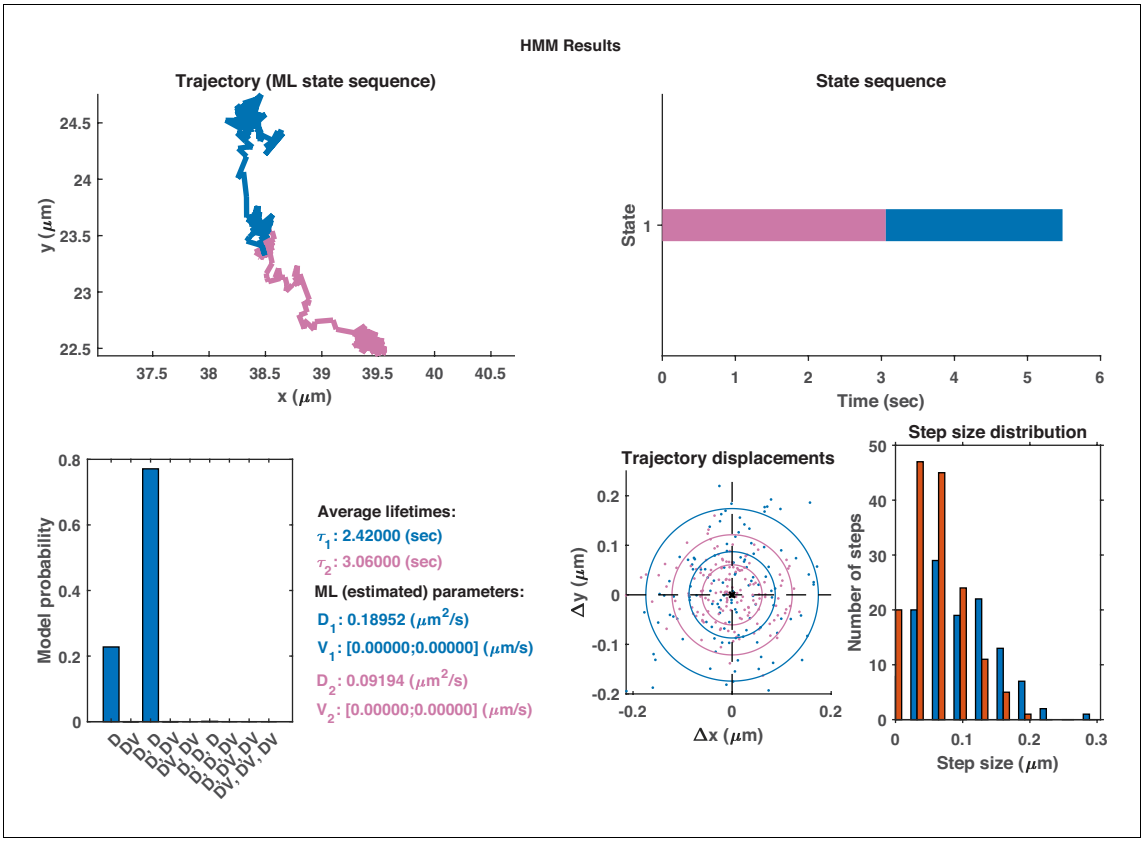
Figure 7. Box and whisker plots of displacements, times and velocities of persistent retrograde, persistent anterograde and anti-persistent segments in experimental trajectories. Any segment with $H>0.55$ was classed as persistent and $H<0.45$ as anti-persistent. These H values were chosen as a precaution against the mean error of the neural network estimation. Each data point within the box and whisker plots are averages of all trajectory segments in a single cell. A total of 65 MRC-5 cells for GFP-Rab5-tagged endosomes, 63 MRC-5 cells for SNX1-GFP-tagged endosomes and 71 MRC-5 cells for lysosomes were analysed with at least 5 to 500 (average 54) anterograde or retrograde segments for each cell.



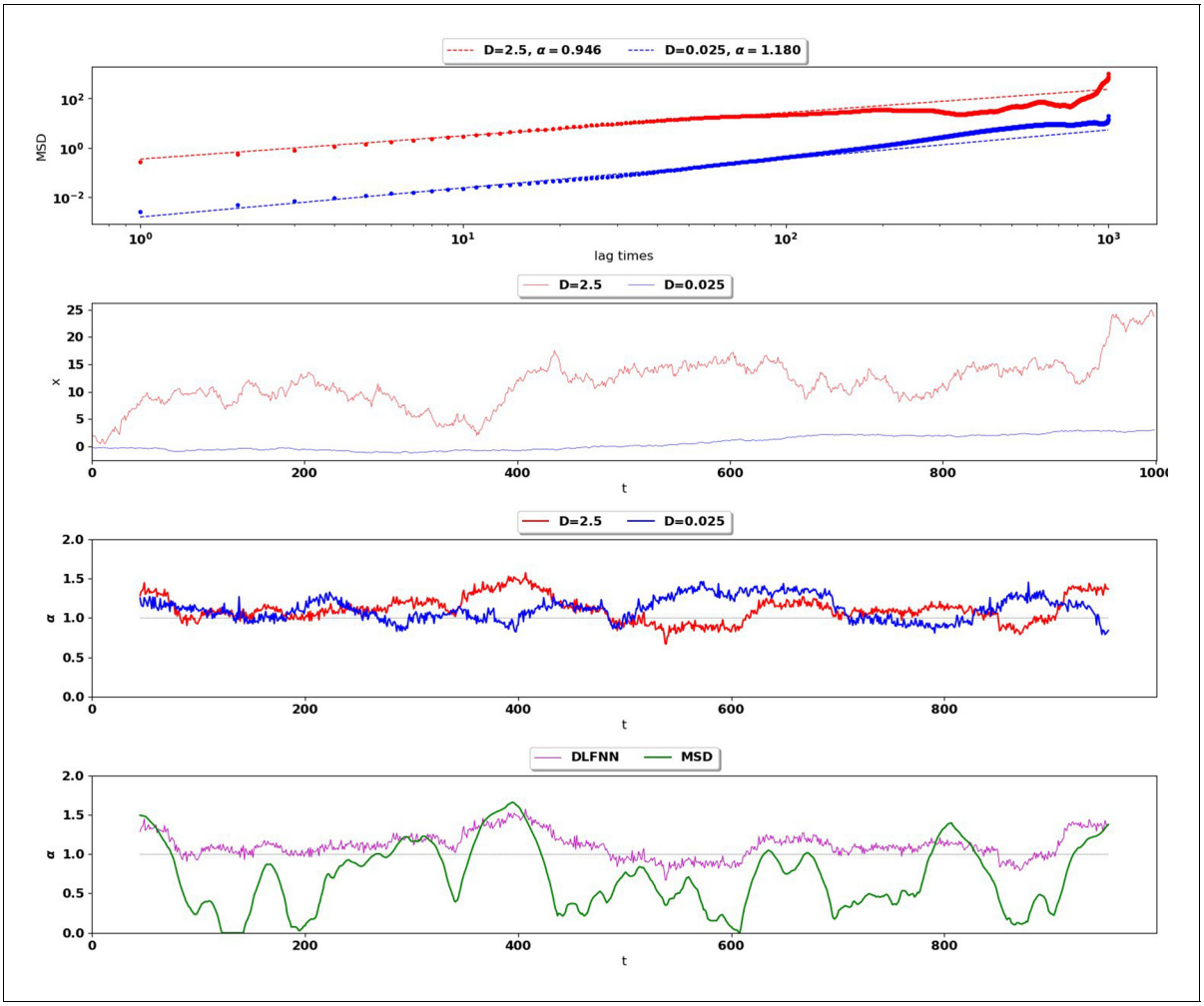
Appendix 1—figure 1. The same single trajectory as in ***Figure 3*** but processed using the HMM-Bayes package described in ***Monnier et al. (2015)***. The plot shows: the original trajectory (*top left*); the inferred state sequence of the most likely model (*top right*); the model probabilities given a maximum of three possible states (*bottom left*); the average lifetime of each state and estimated parameters of the most likely model (*bottom center left*), which in this case is two different diffusive states; individual increment displacements (*bottom center right*); and the step size distribution of those increments classed into the two different states.



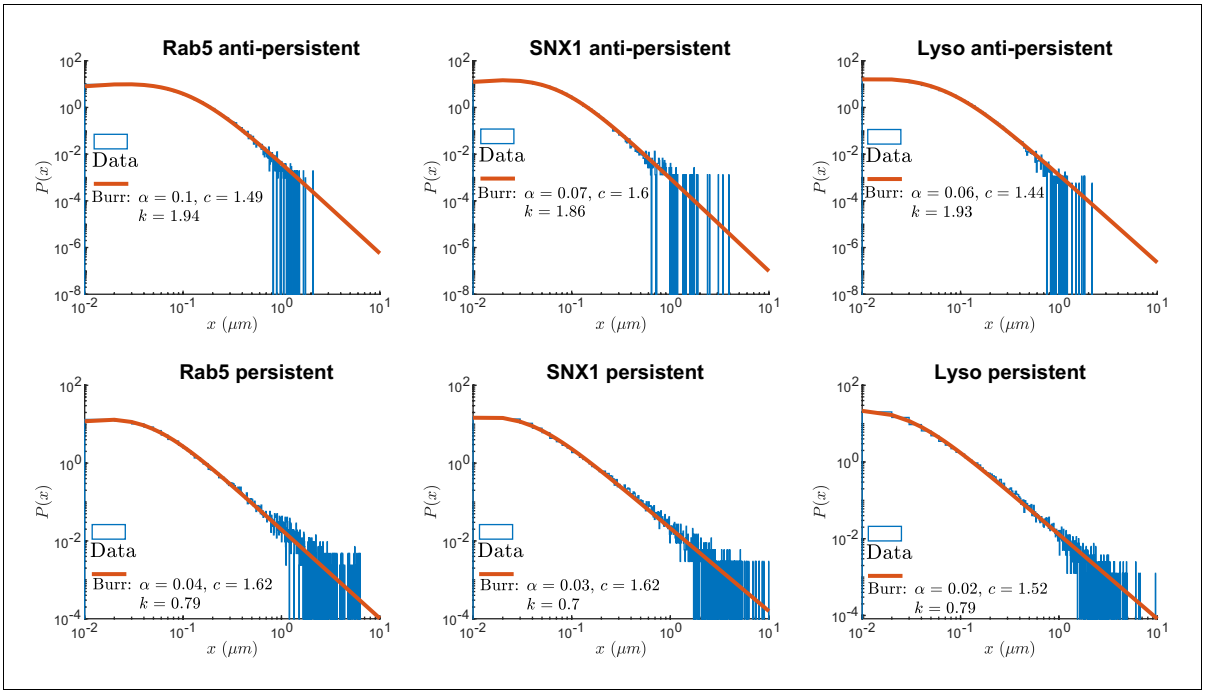
Appendix 1—figure 2. *Top*: Plot of displacement from a single GFP-SNX1 endosome trajectory in an MRC-5 cell (blue). Shaded areas show persistent ($0.55<H<1$ in green) and anti-persistent ($0<H<0.45$ in magenta) behaviour. *Middle*: A 15 point moving window DLFNN exponent estimate for the trajectory (black) with a line (dashed) marking diffusion $H = 0.5$ and lines (dotted) marking the confidence bounds $H = 0.55$ and 0.45. *Bottom*: Plot of instantaneous and moving (15 point) window velocity. *Right*: Plot of the trajectory of a GFP-SNX1 endosome in an MRC-5 cell with start and finish positions, and persistent (green) and anti-persistent (magenta) segments indicated.



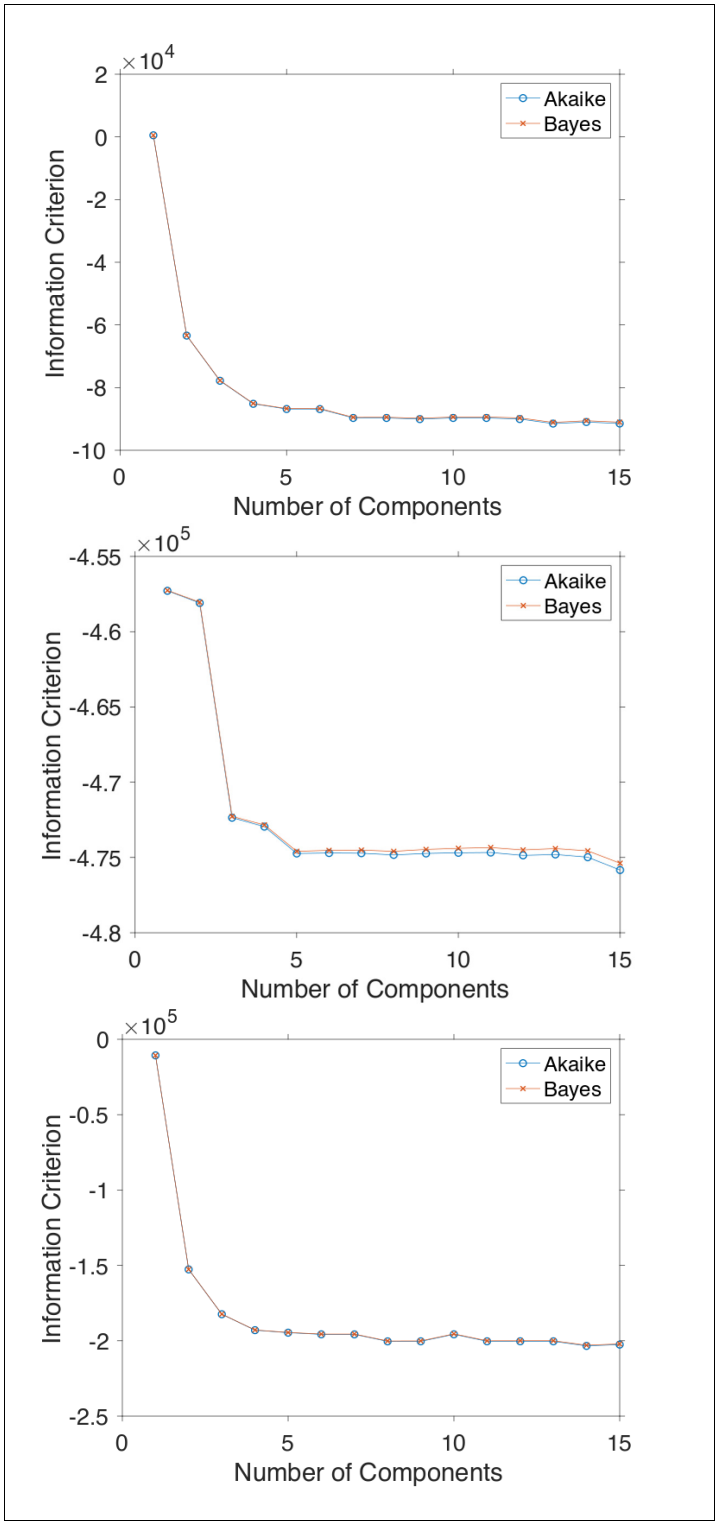
Appendix 1—figure 3. The same HMM-Bayes analysis as shown in ***Appendix 1—figure 1*** applied to the trajectory in ***Appendix 1—figure 2***.



Appendix 2—figure 1. Top: MSD (points) and power-law fits (dashed) for two different Brownian trajectories containing 1000 data points with diffusion coefficient 2.5 (red) and 0.025 (blue). The Hurst exponent should be $\alpha = 2H = 1$ for both trajectories. Second Row: Simulation of the two Brownian trajectories with diffusion co-efficient 2.5 (red) and 0.025 (blue). Third Row: Local Hurst exponent estimates given by the DLFNN for the two different trajectories using a 90 point window. The averages of DLFNN Hurst exponent estimates are $\alpha = 2H = 1.110$ (red) and $\alpha = 2H = 1.114$ (blue). Bottom: Local Hurst exponent estimates of the $D = 2.5$ track given by DLFNN and MSDs using a 90 point window. The average of DLFNN Hurst exponent estimates is $\alpha = 2H = 1.110$ and the average of MSD estimates is $\alpha = 2H = 0.937$.



Appendix 3—figure 1. Normalized histograms (blue) and corresponding maximum likelihood estimation for Burr distributions (line) of segment displacements from lysosome and endosome experimental trajectories segmented using DLFNN. Parameter estimates are shown the legend.



Appendix 4—figure 1. The Akaike and Bayes information criterion against number of components in the Gaussian mixture model shown in ***Figure 5*** for GFP-Rab5 tagged endosomes (top), SNX1-GFP tagged endosomes (middle) and lysosomes (bottom).