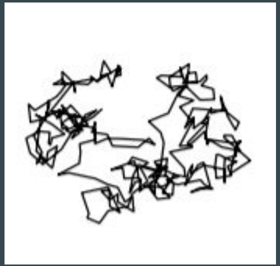
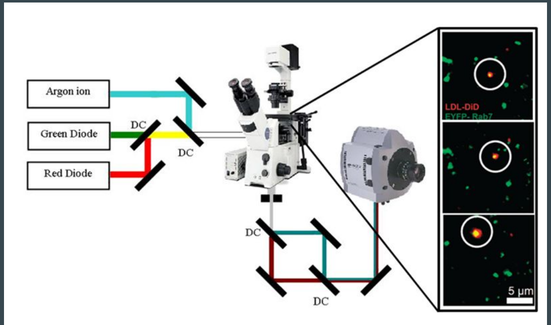
A first-pass statistical dashboard for categorizing diverse particle movement patterns

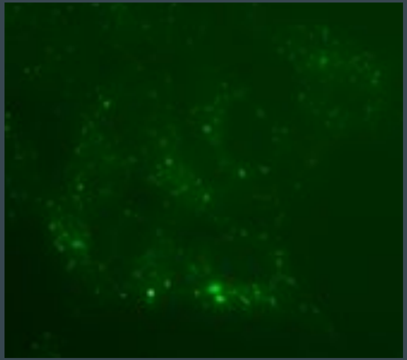
Riley Juenemann
Tulane University

Advisors:
Scott McKinley, Mathematics, Tulane University
Christine Payne, Materials Science and Mechanical Engineering, Duke University



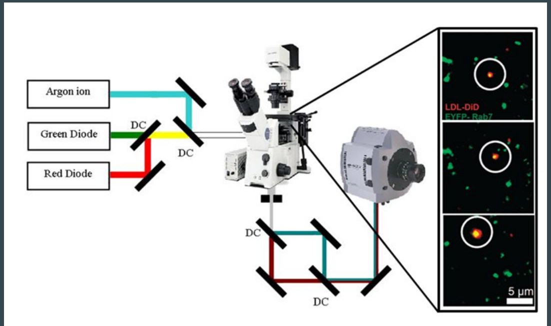
Problem: Dynamics within cells are not well understood

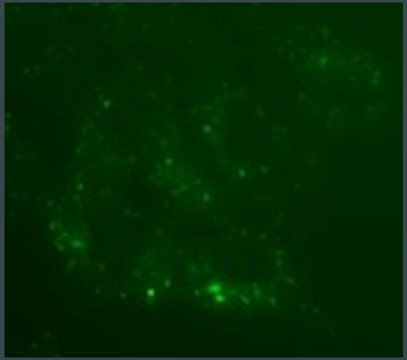


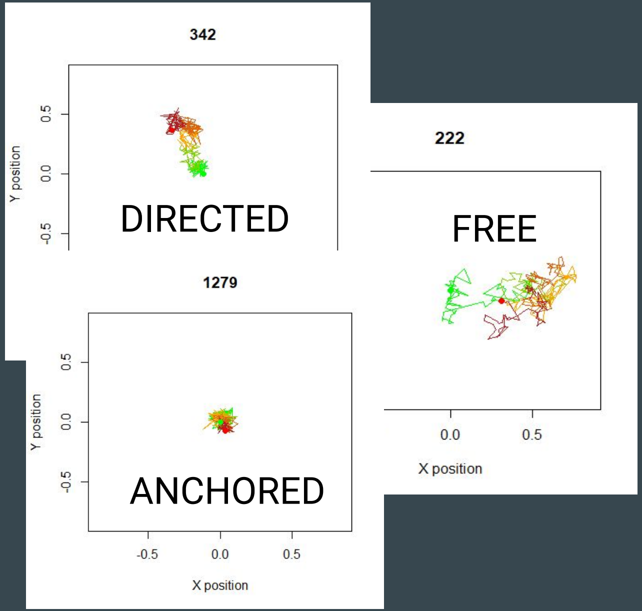


Problem: Dynamics within cells are not well understood

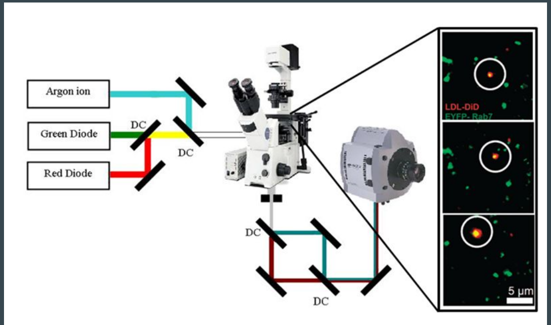
Goal: Automate categorization of particle movement in cells

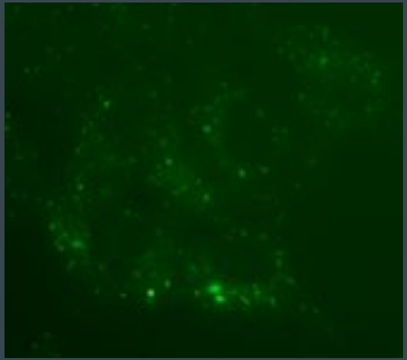




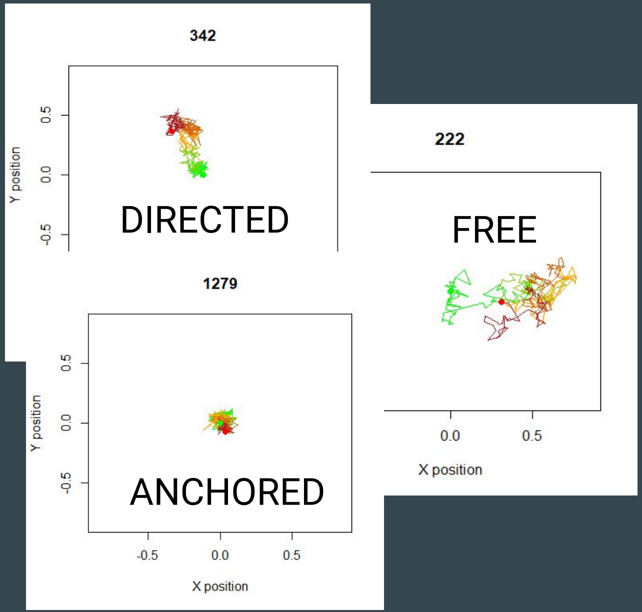


Problem: Dynamics within cells are not well understood

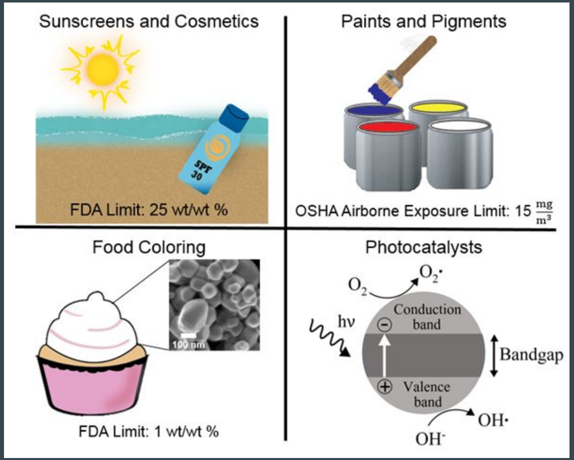




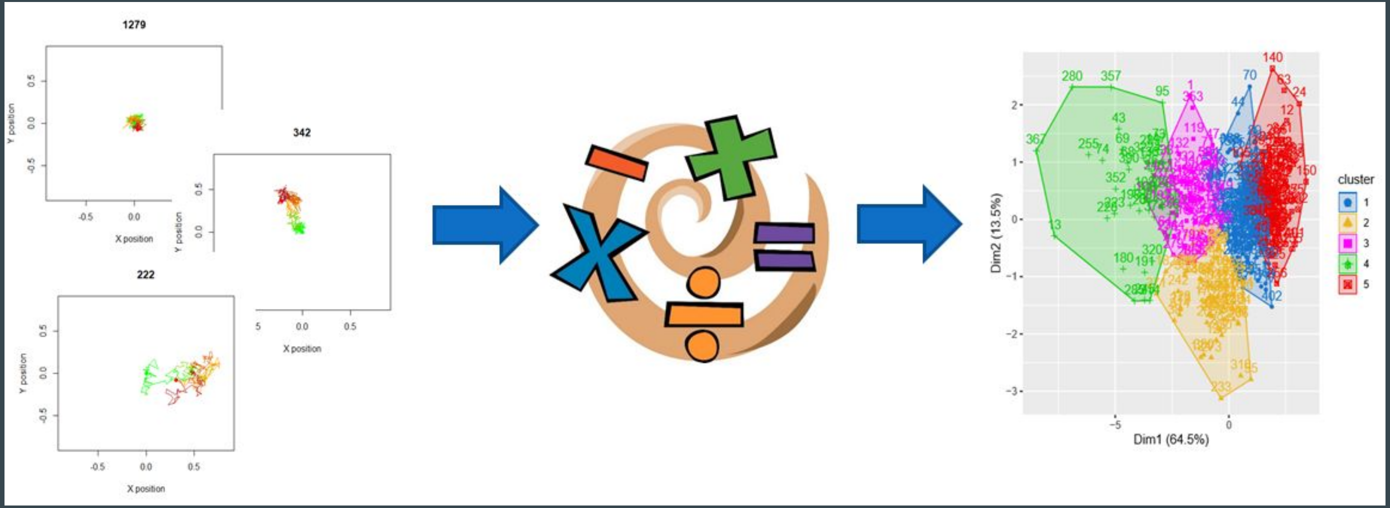
Goal: Automate categorization of particle movement in cells



Application: Titanium Dioxide Nanoparticles in Human Lung Cells



How can we use math to categorize diverse movement types?



Start with data we "know" -> simulated paths

Zero Mass (Overdamped) Langevin Dynamics

$$\gamma dX(t) = -\kappa(X(t) - Z(t))dt + \sqrt{2k_B T\gamma}dW(t)$$

Derived from $F = ma$

$$ma = F_{drag} + F_{anchor} + F_{thermal}$$

$$m\ddot{X} = -\gamma\dot{X} - \kappa(X - Z) + \sqrt{2D}\dot{W}$$

Assume $m \approx 0$ and bring $-\gamma\dot{X}$ to the left

Employ stochastic differential equation notation

$$dX(t) = -\frac{\kappa}{\gamma}(X(t) - Z(t))dt + \sqrt{2D}dW(t)$$

X - particle position

Z - anchor location

κ - spring constant (assuming Hooke's Law)

γ - drag coefficient

$D = \frac{k_B T}{\gamma}$ - diffusivity

k_B - Boltzman's Constant

T - temperature

W - Brownian motion

We can use this equation to simulate the easy categories!

$$dX(t) = -\frac{\kappa}{\gamma}(X(t) - Z(t))dt + \sqrt{2D}dW(t)$$

X - particle position
Z - anchor location
κ - spring constant
(assuming Hooke's Law)
γ - drag coefficient
D - diffusivity
W - Brownian motion

Scenarios:

(1) $\kappa = 0$ - Free Diffusion

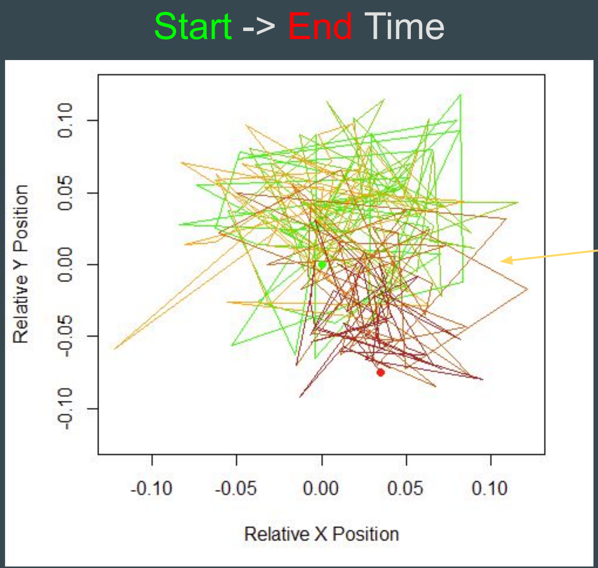
Simulation: cumulative sum of Gaussian random variables

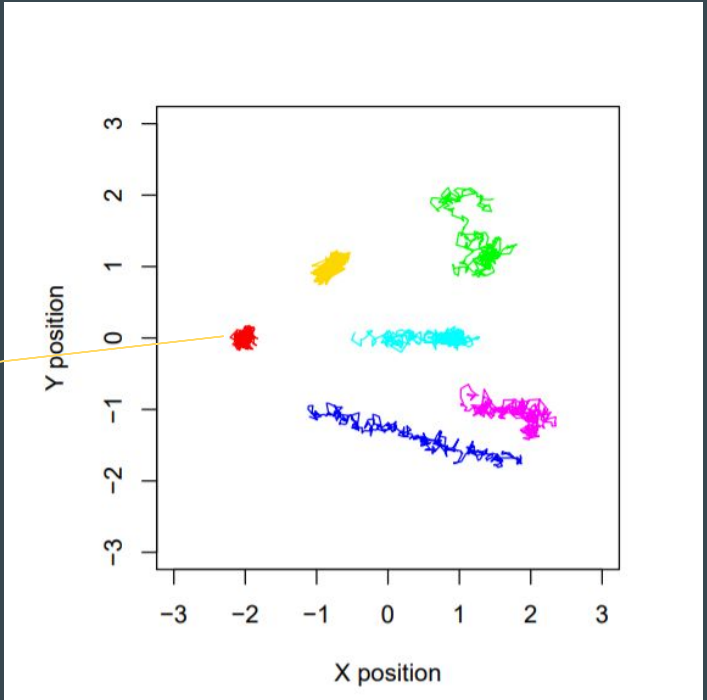
(2) $Z(t) = z$ is fixed - Stuck Diffusion

(3) $Z(t) = mt + b$ - Directed Transport

Scenarios 2 and 3 allow for an explicit solution that we can use for simulation.

Simulated data types





Free Diffusion Anchored Diffusion Directed Transport

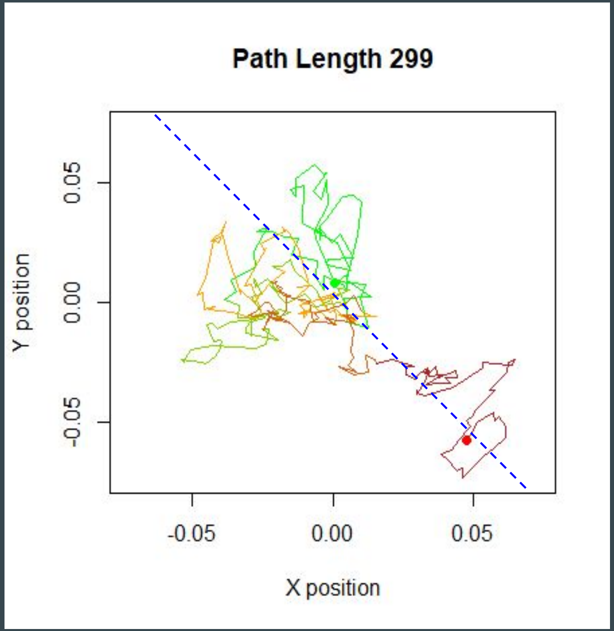
Tracker Error Subdiffusion Skating Diffusion

Can choose parameters to match live cell data:

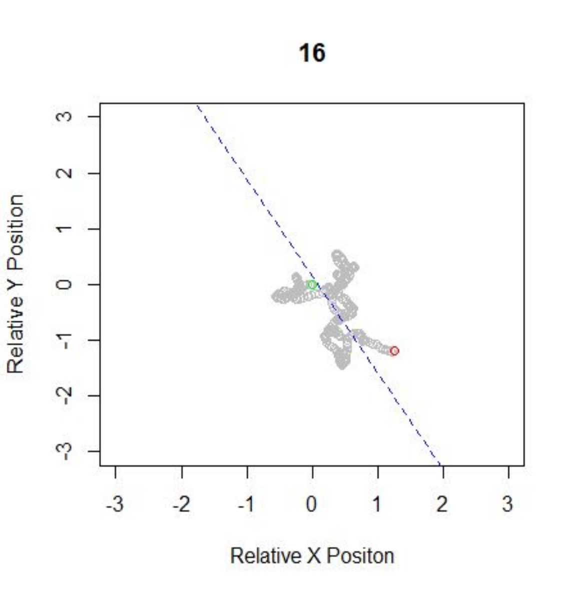
Diffusivity = 0.002 , Velocity = 0.4/300, κ = 1

Projecting onto major and minor axis

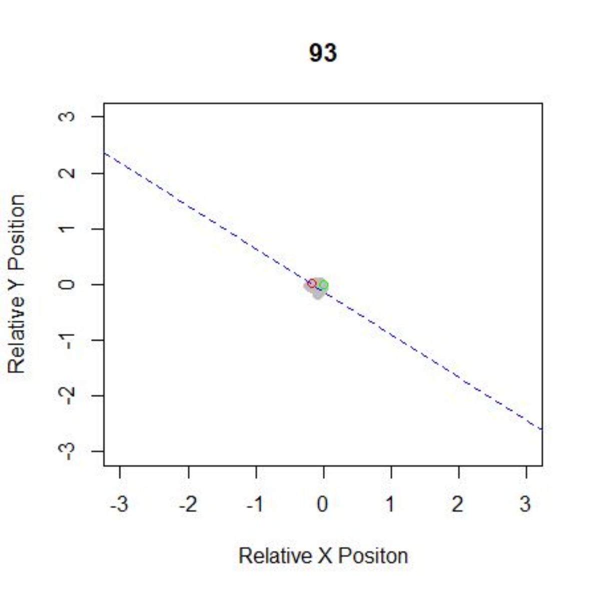
- Project all the points onto lines at various angles between 0 and π
- Find the long direction (line at angle theta with greatest range of points)
  - This axis is the major direction
- Find the transverse direction
  - This axis is the minor direction
- Take the extent (standard deviation) for both directions
  - Major direction -> SD_{major}
  - Minor direction -> SD_{minor}



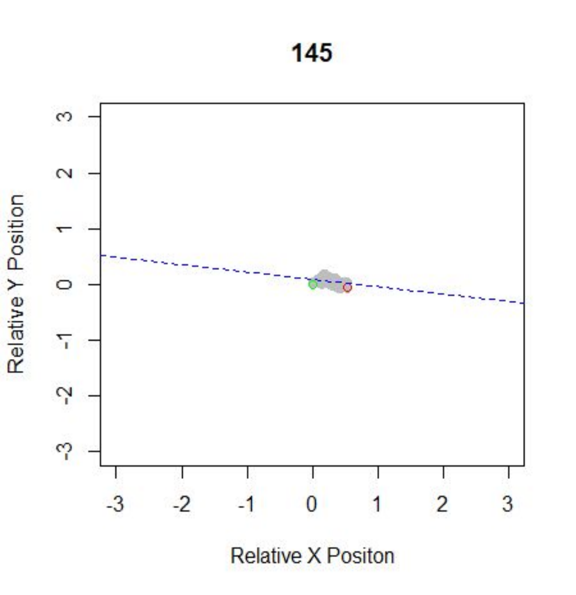
How does the standard deviation differ for different types?



Freely Diffusing



Anchored Diffusing



Directed Transport

We can use these to isolate non-diffusive patterns

Freely Diffusing

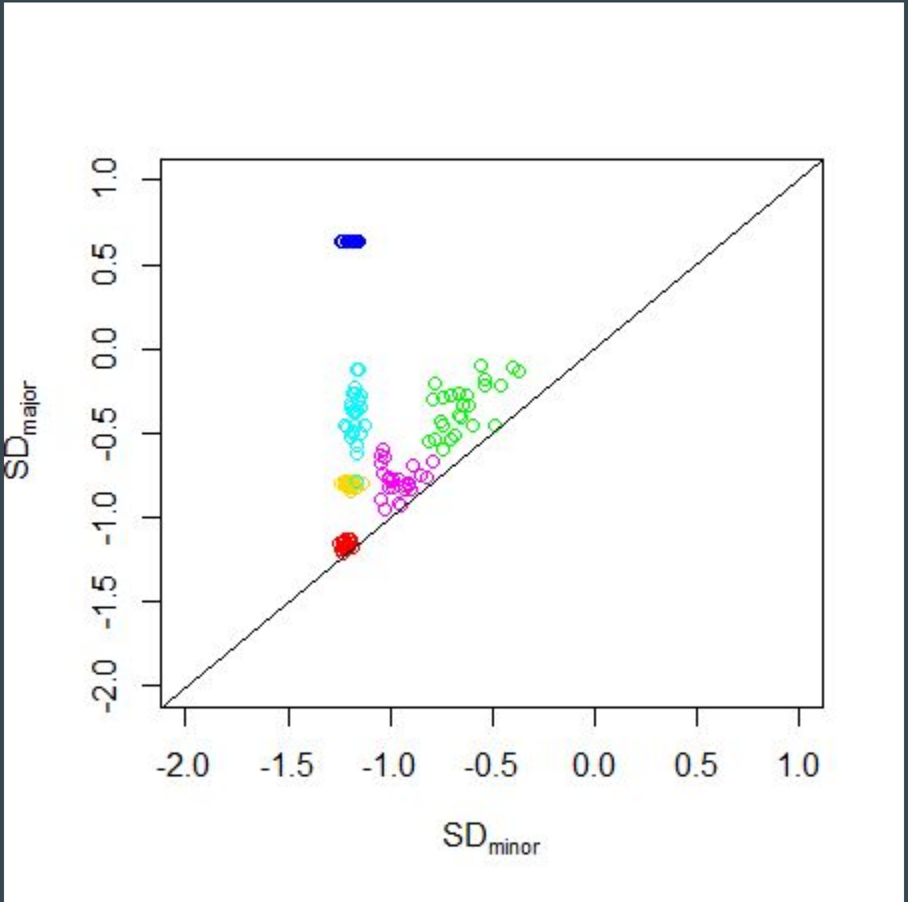
Stuck Diffusing

Directed Transport

Tracker Error

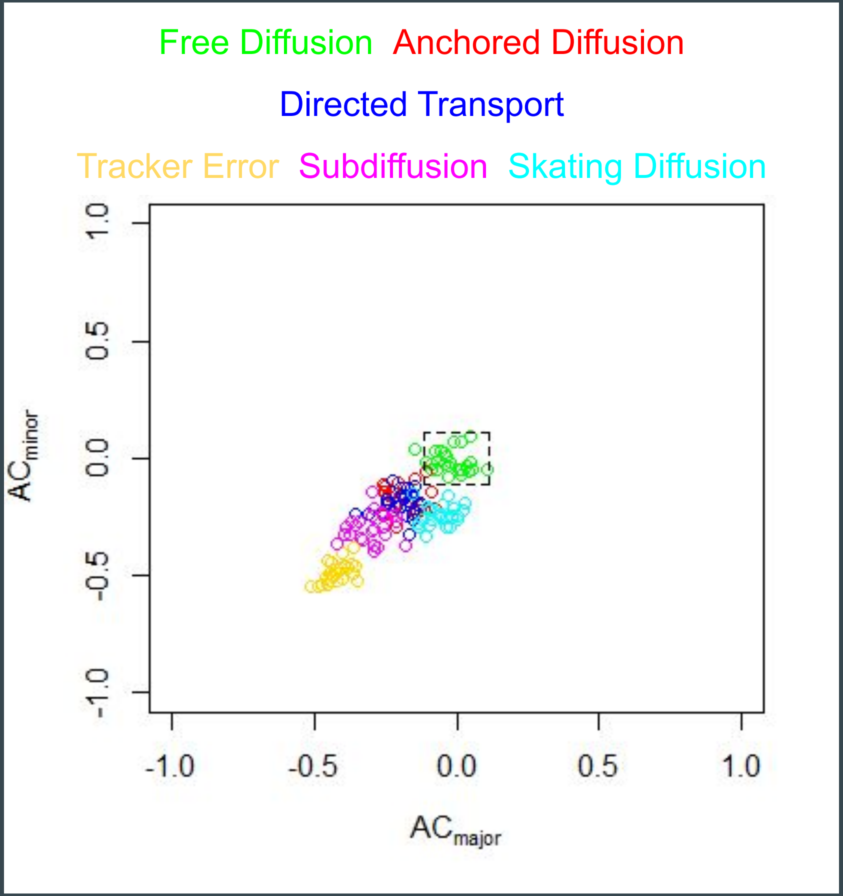
Subdiffusing

Skating



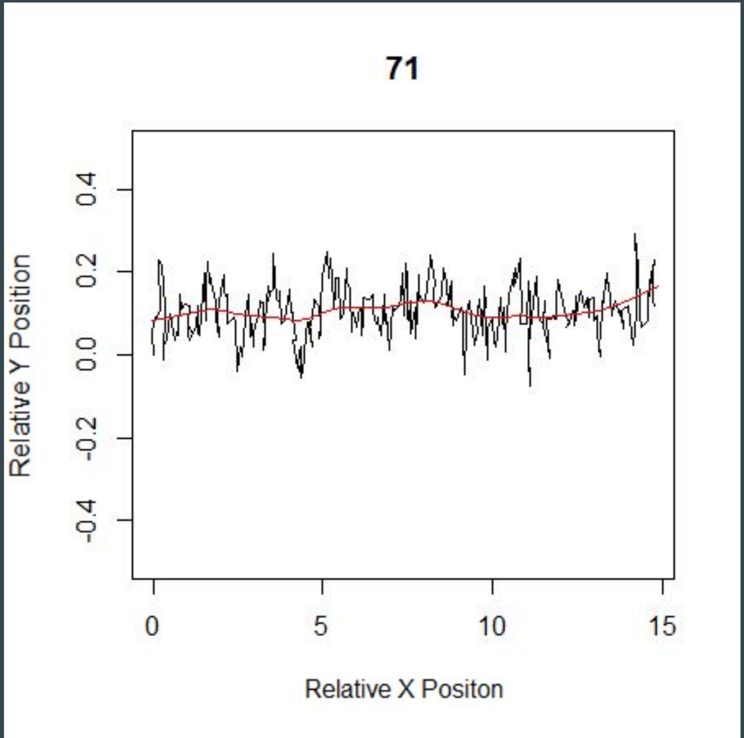
Free diffusion has independent increments

- Test for independence of increments with Autocorrelation Function (ACF) time lag 1
- AC_{major} = correlation values for consecutive increments in the major direction
- AC_{minor} = correlation values for consecutive increments in the minor direction



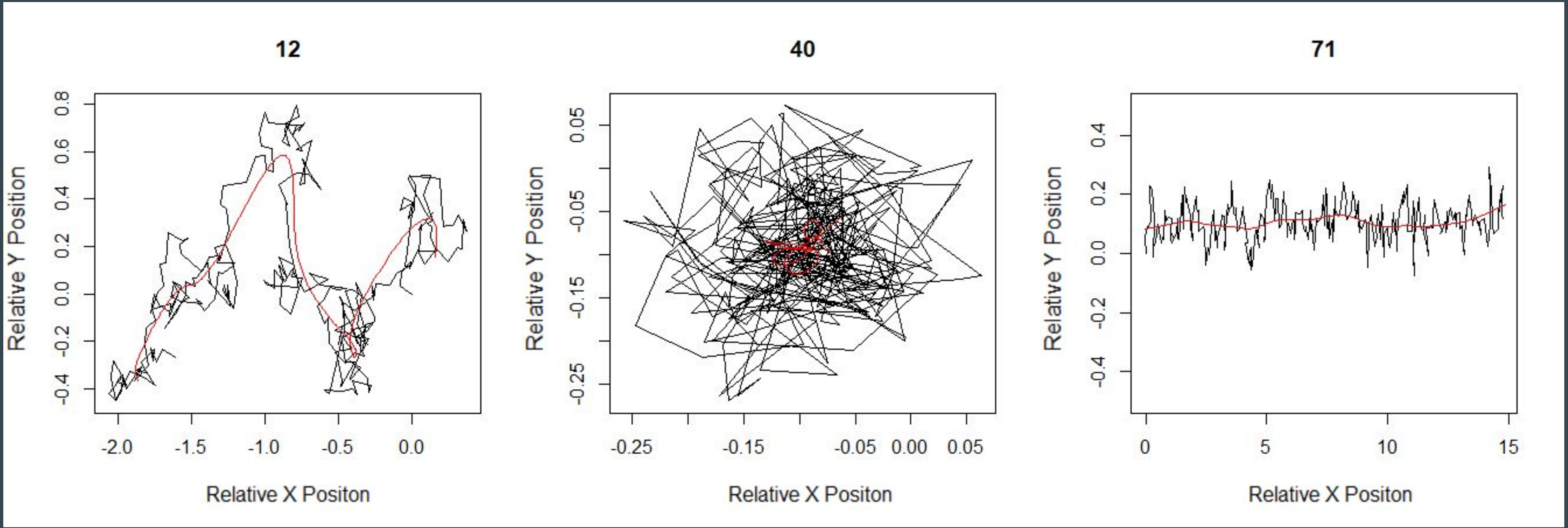
Estimating variance around anchor

- Estimate anchor path from parametrizing splines
- Compute distance to anchor from observed particle location at each time step
- Take the average of the squared distances for each path (δ_{anchor})



Splines on easy simulated data

Units: microns (µm)



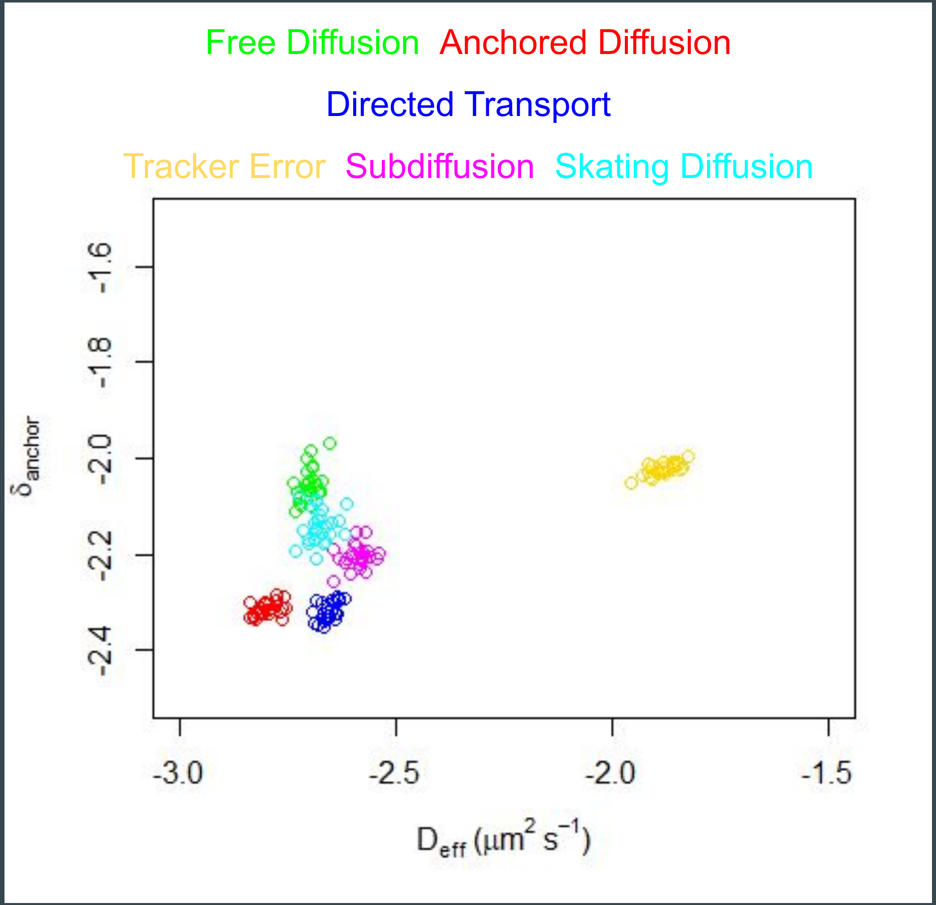
Free Diffusion

Anchored Diffusion

Directed Transport

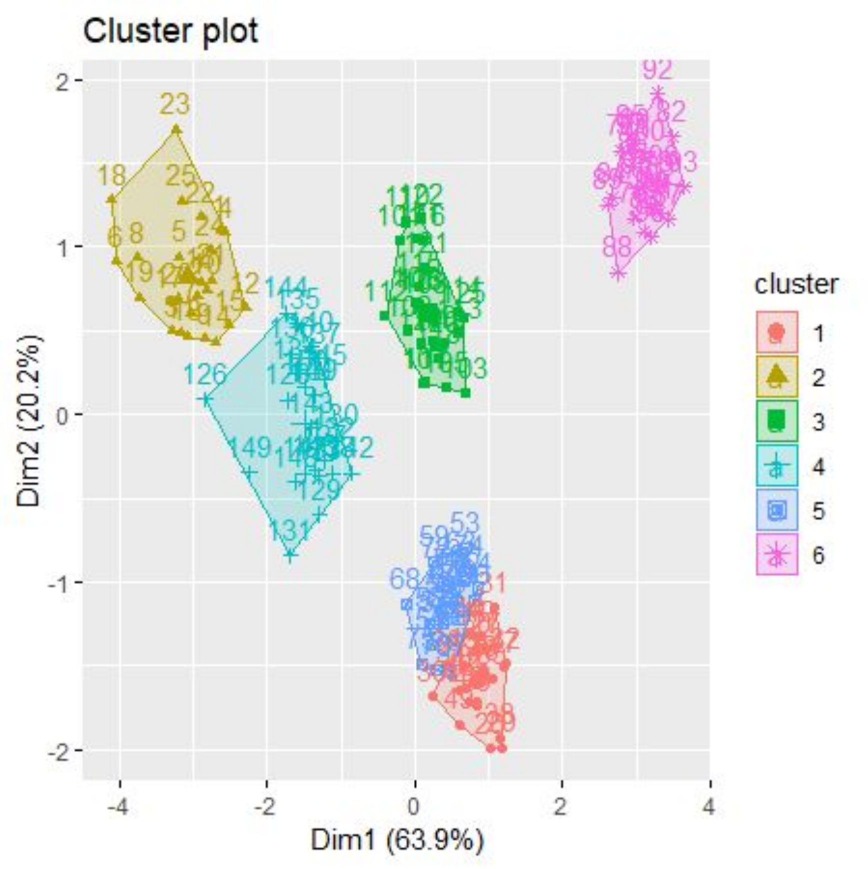
Diffusivity and anchor distance separate tracker error

- D_{eff} comes from using Maximum Likelihood Estimation (MLE)
  - X and Y increment processes assuming the particles are freely diffusing
- δ_{anchor} is distance from splines



K-means clustering allows us to see in 6 dimensions!

- Unsupervised machine learning algorithm
- Partitions a data set into k clusters
  - k is specified by the user
- Input:
  - 25 of each simulation type
  - AC_{maj}, AC_{min}, D_{eff}, SD_{major}, SD_{minor}, δ_{anchor}
- All correctly categorized!



Simulated data has validated the dashboard, what's next?

Given data sets, we now have an automated categorization system that will help us identify cells with more directed transport and cells with more diffusive transport.

We have applied this dashboard on datasets from our collaborators.

Simulated data has validated the dashboard, what's next?

Given data sets, we now have an automated categorization system that will help us identify cells with more directed transport and cells with more diffusive transport.

We have applied this dashboard on datasets from our collaborators.

Current Extensions:

- (Computer Science Capstone) Using supervised machine learning (support vector machine) instead of k-means
- (Math Honors Thesis) Model selection for diffusive movement patterns
- What is the ideal number and frequency of observations for fluorescence microscopy experiments?

Questions?

Email: rjuenemann@tulane.edu
LinkedIn: https://www.linkedin.com/in/rjuenemann/
Twitter: @rileyj2017

I would like to acknowledge:

- NCUWM and the University of Nebraska–Lincoln
- Tulane University and Prof. Scott McKinley's Stochastics Lab
- National Institutes of Health (NIH) R01GM122082-01 - "Bridging Understanding of Motor-Cargo Transport from Artificial to Cellular Systems"
- Duke University and Prof. Christine Payne's Lab
  - Summer experimental training funded by National Science Foundation (NSF) through the Research Experiences for Undergraduates (REU) program