

# **High-Performance Software Pipeline for Single Molecule Tracking Data**

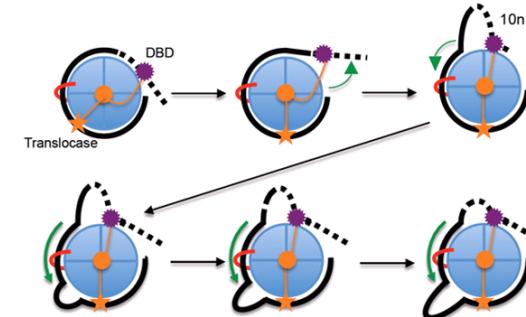
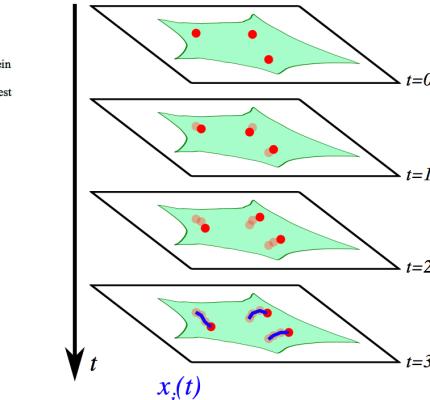
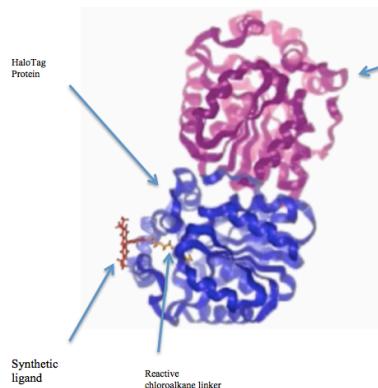


**Dec. 4, 2017**

**Sun Jay Yoo – Wu Lab / HHMI**

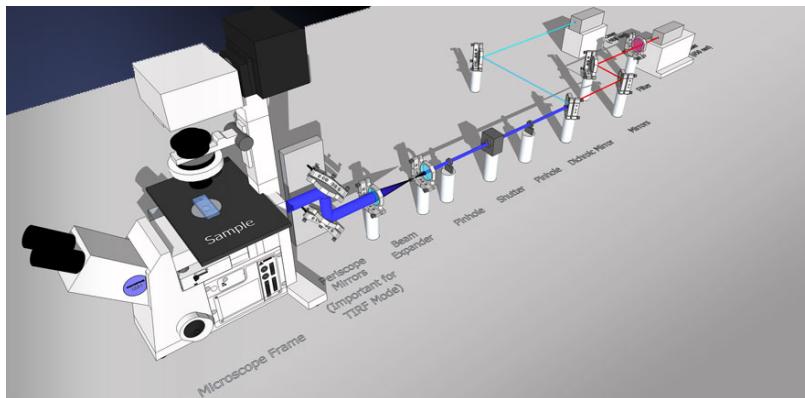
# General Overview

- Fluorescent biomarkers and tags
- Super-resolution single molecule imaging
- Mechanistic and conceptual models for underlying cell processes



# General Overview

- Fluorescent biomarkers and tags
- Super-resolution single molecule imaging



- Positions on nanometer scale (~30nm)
- Movement in millisecond scale (~10ms)

## General Overview

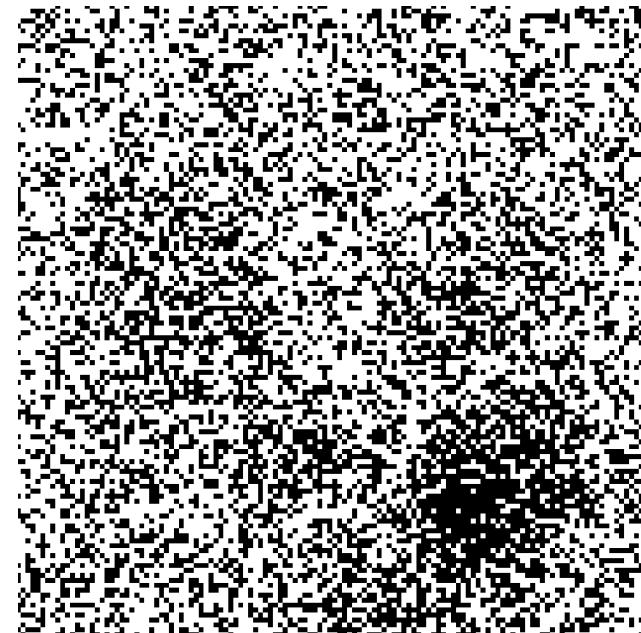
- Super-resolution  
single molecule  
imaging



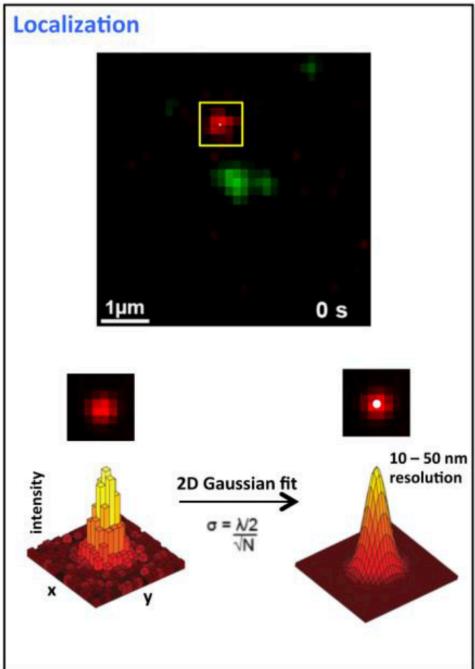
# General Overview

- Super-resolution single molecule imaging

*With thresholding*



# General Overview



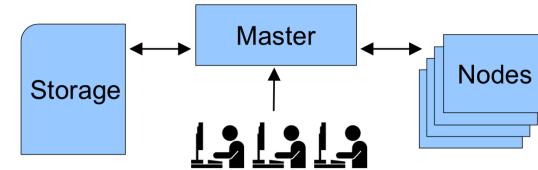
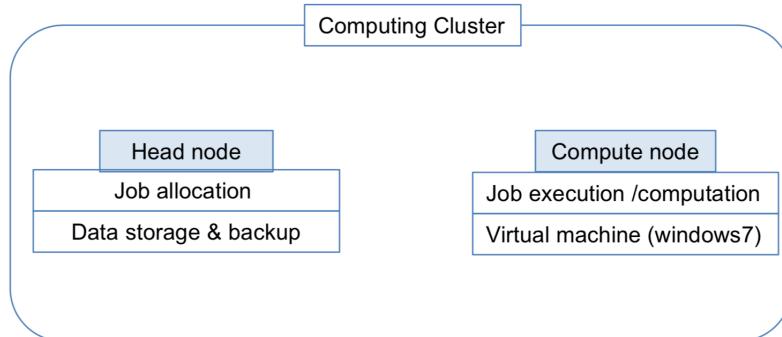
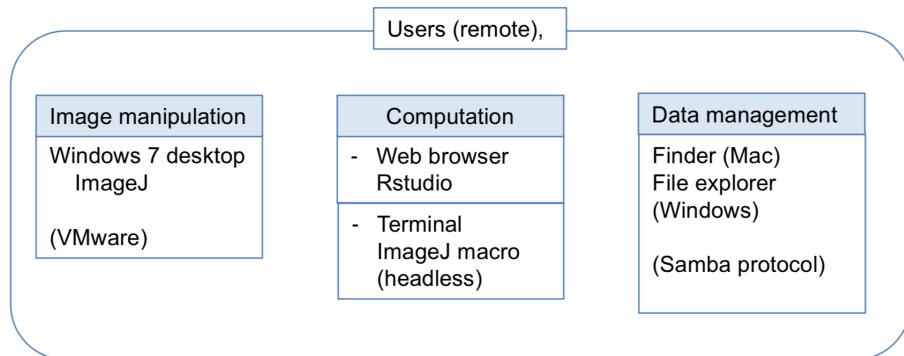
## ■ Super-resolution single molecule imaging

```
>> tracks{1}  
ans =  
  
struct with fields:  
  
RefinedCooY: [11.7849 44.1734 75.9684 73.8330 60.6965]  
RefinedCooX: [38.8411 60.0523 64.5863 74.8155 82.8679]  
RefinedCooZ: [1.0001 1.0001 1.0000 1.0001 1.0001]  
Intensity: [126.4305 124.8418 92.7616 129.8724 230.6995]  
GoodnessFit: [6.4093e-06 7.8323e-06 2.4948e-06 3.8869e-06 3.8296e-06]  
Successor: [0 2 3 5 6]  
Predecessor: [0 0 0 0 0]
```

# General Overview

- Super-resolution  
single molecule  
imaging
  - Mechanistic and  
conceptual models  
for underlying cell  
processes
- 
- 

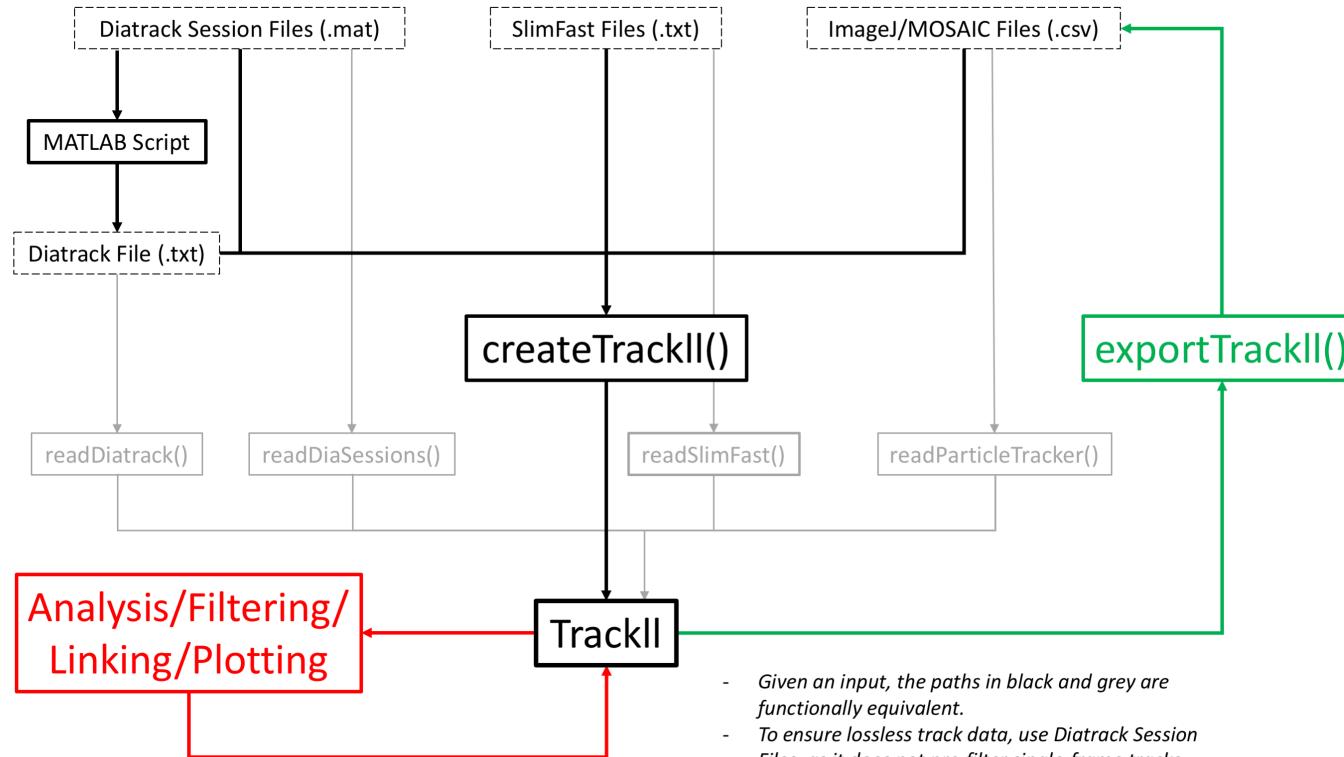
# Preliminary Setup



■ **High performance  
compute cluster**

■ **File server and  
remote workstation**

# The Software



**~ 8000 lines of code**

- Given an input, the paths in black and grey are functionally equivalent.
- To ensure lossless track data, use Diattrack Session Files, as it does not pre-filter single-frame tracks.

# 1

## Reading Single Molecule Tracks



# Reading Single Molecule Tracks

```
>> tracks{1}
```

```
ans =
```

```
struct with fields:
```

```
  RefinedCooY: [11.7849 44.1734 75.9684 73.8330 60.6965]
  RefinedCooX: [38.8411 60.0523 64.5863 74.8155 82.8679]
  RefinedCooZ: [1.0001 1.0001 1.0000 1.0001 1.0001]
  Intensity: [126.4305 124.8418 92.7616 129.8724 230.6995]
  GoodnessFit: [6.4093e-06 7.8323e-06 2.4948e-06 3.8869e-06 3.8296e-06]
  Successor: [0 2 3 5 6]
  Predecessor: [0 0 0 0 0]
```

```
>> tracks{2}
```

```
ans =
```

```
struct with fields:
```

```
  RefinedCooY: [37.3711 43.6777 76.0090 70.4019 74.6256 60.8150]
  RefinedCooX: [29.0089 59.0154 64.7354 73.0693 75.0821 82.8030]
  RefinedCooZ: [1.0001 1.0000 1.0001 1.0000 1.0000 1.0001]
  Intensity: [84.8280 106.2075 144.0688 133.8777 128.6550 232.1278]
  GoodnessFit: [5.7017e-06 3.8591e-06 4.1910e-06 2.1817e-06 5.2488e-06]
  Predecessor: [0 2 3 0 4 5]
  Successor: [0 1 2 3 0 4]
```

```
>> tracks{3}
```

```
ans =
```

```
struct with fields:
```

```
  RefinedCooY: [44.8422 74.7096 72.0128 60.8010]
  RefinedCooX: [58.8154 66.7509 73.2884 82.8456]
  RefinedCooZ: [1.0000 1.0000 1.0001 1.0001]
  Intensity: [91.1093 120.7932 170.1534 246.6156]
  GoodnessFit: [6.5521e-07 2.5930e-06 3.1175e-06 7.3064e-08]
  Predecessor: [2 3 4 6]
  Successor: [1 0 2 3]
```

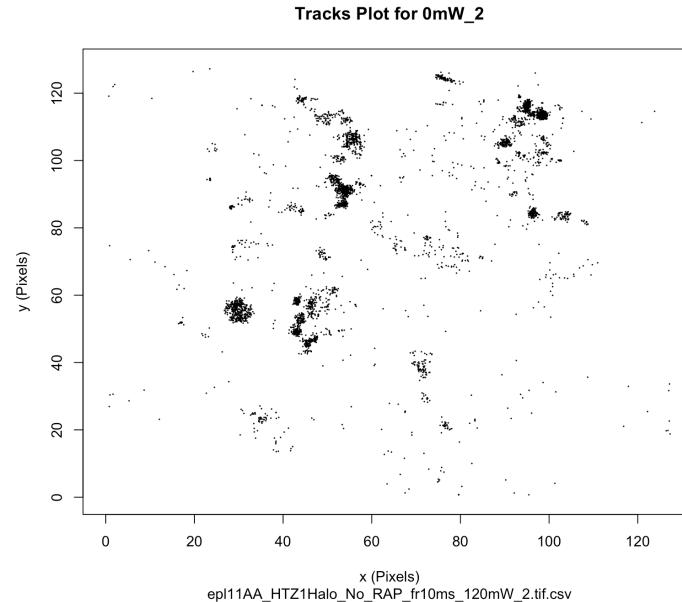
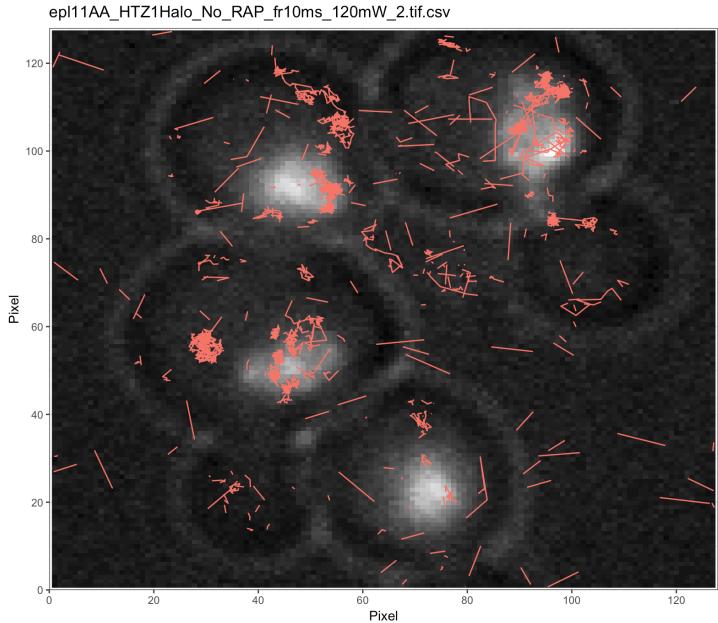
```
> track11 <- readDiaSessions()
Reading Diatrack session file:
HT21Halo_frl0ms_120mW_5.mat
...
Session file read and
processed.
```

```
> track11[1:5]
```

```
$`OmW_5.1.4.2`  
x y z frame  
1 62.97 95.19 1 1  
2 63.25 94.90 1 2  
3 63.77 94.81 1 3  
4 62.99 94.77 1 4  
$`OmW_5.1.13.3`  
x y z frame  
1 90.26 70.80 1 1  
2 90.29 70.89 1 2  
3 90.64 71.04 1 3  
4 90.21 72.91 1 4  
5 89.98 73.37 1 5  
4 89.73 73.24 1 6  
5 89.31 84.04 1 5  
6 89.20 72.09 1 7  
6 89.70 71.09 1 8  
7 89.38 70.96 1 9  
8 89.69 72.88 1 10  
9 89.77 72.91 1 11  
$`OmW_5.3.2.4`  
x y z frame  
1 45.13 64.20 1 3  
2 45.82 64.13 1 4  
$`OmW_5.6.10.5`  
x y z frame  
1 45.28 64.08 1 6  
2 45.05 64.33 1 7  
3 45.81 64.24 1 8  
4 45.33 64.15 1 9  
5 45.80 63.97 1 10  
6 45.32 64.06 1 11  
7 46.01 64.67 1 12  
8 45.15 64.80 1 13  
9 45.08 64.91 1 14  
10 45.14 64.90 1 15
```



# Reading Single Molecule Tracks



# 2

## Refining Data



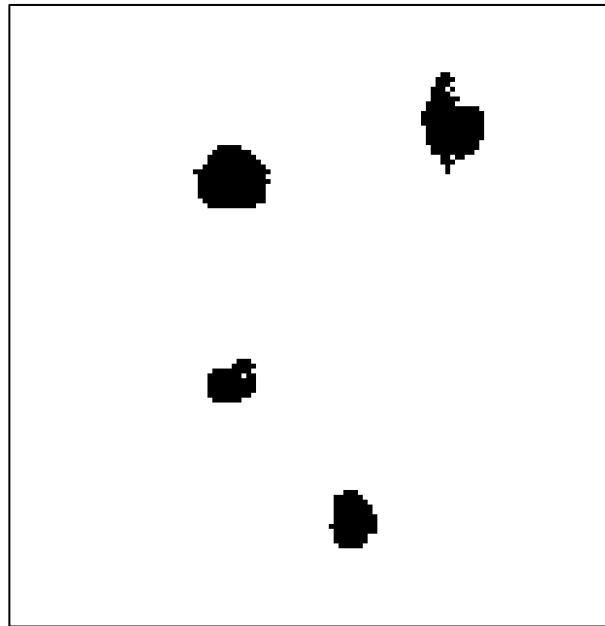
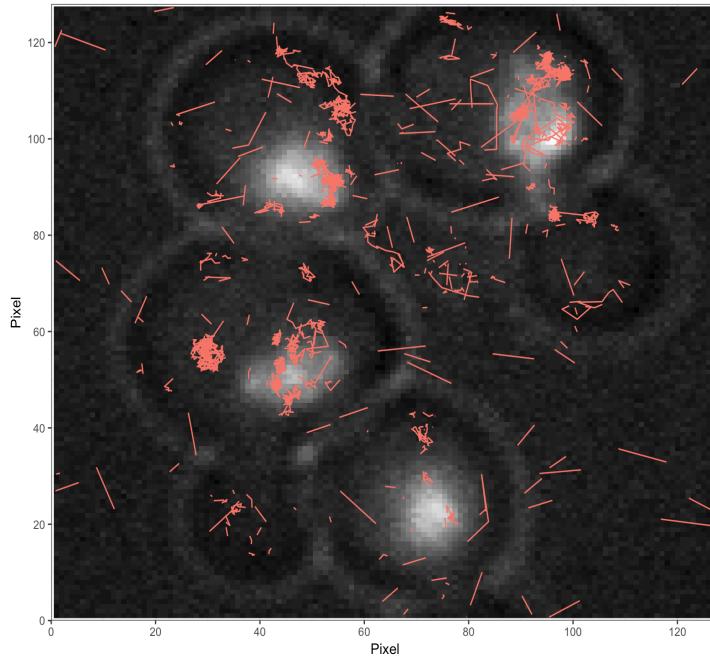
## Refining Data

- **Filtering**
- **Linking**
- **Trimming**
- **Merging**
- **Masking**
- **Masking \***



## Refining Data

epI11AA-HTZ1Halo\_No\_RAP\_fr10ms\_120mW\_2.tif.csv

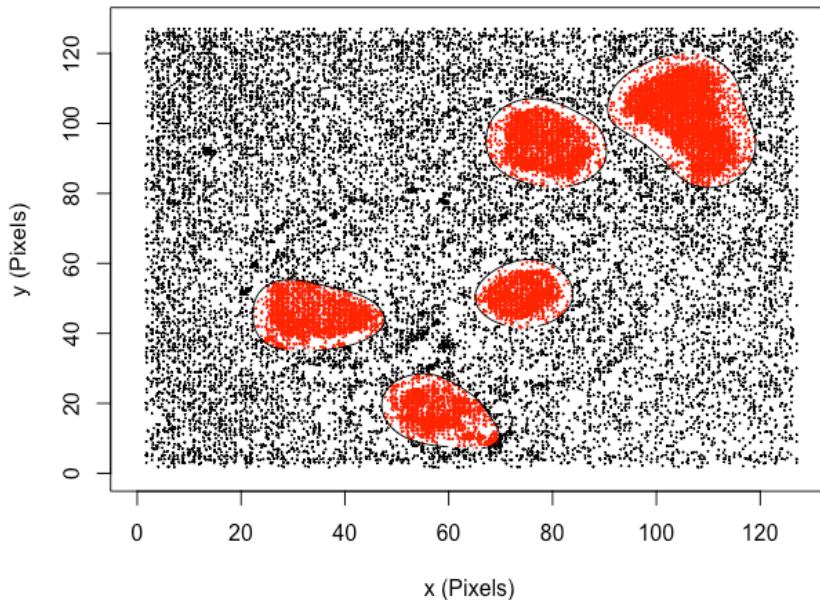


**Manually masking**



## Refining Data

0mW\_6 Mask with Kernel Density Probability ( $p$ ) of 0.5

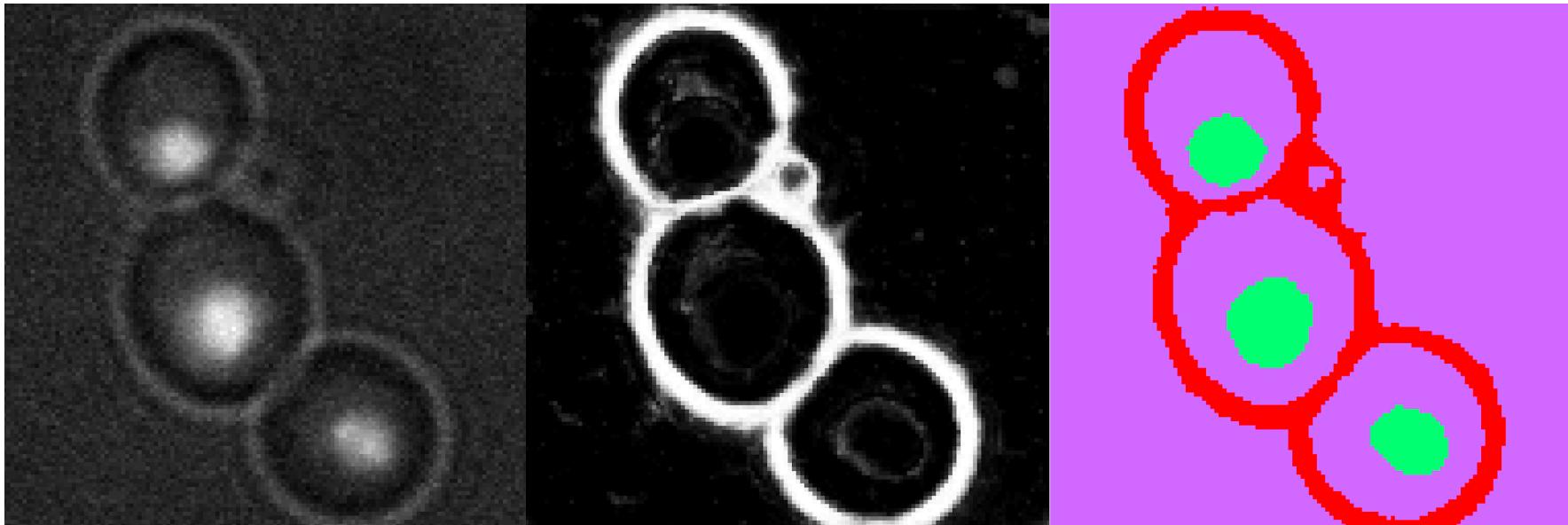


***Masking by kernel  
density***



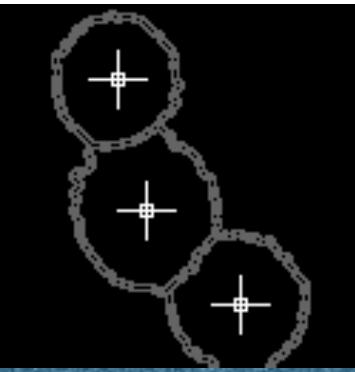
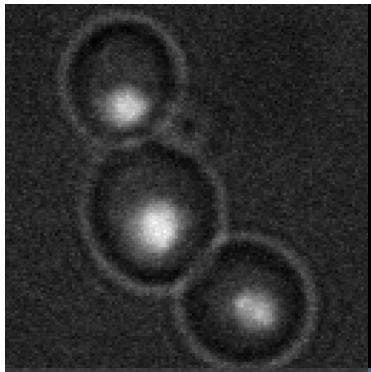
## Refining Data

***Masking by segmentation  
and extraction***

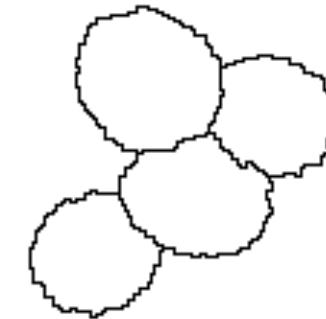
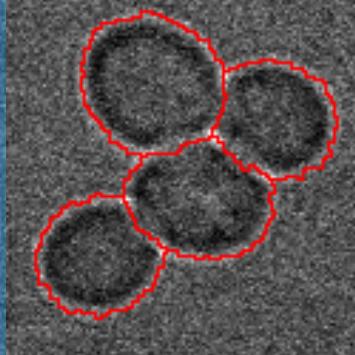
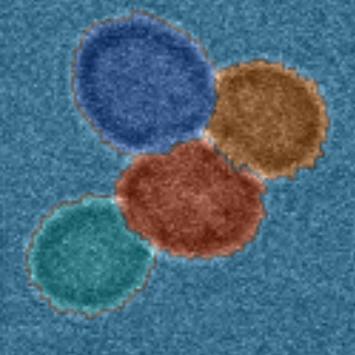
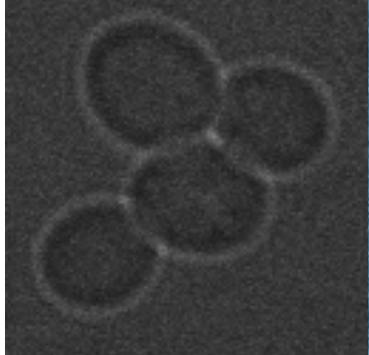




## Refining Data



***Circle detection using the Hough transform***



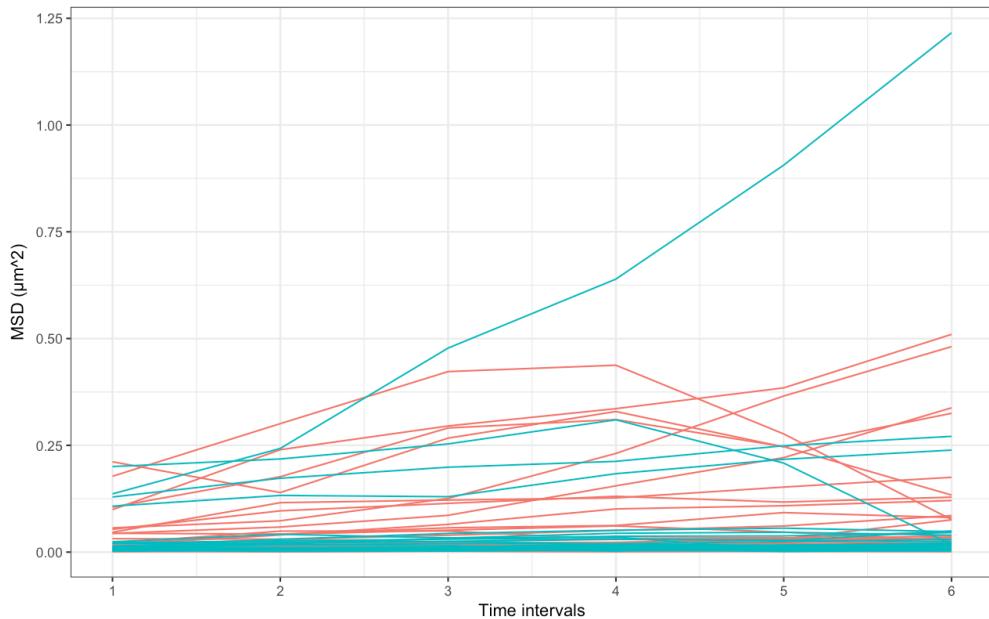
***Simple cell segmentation***

# 3

## Analyzing Data



# Analyzing Data



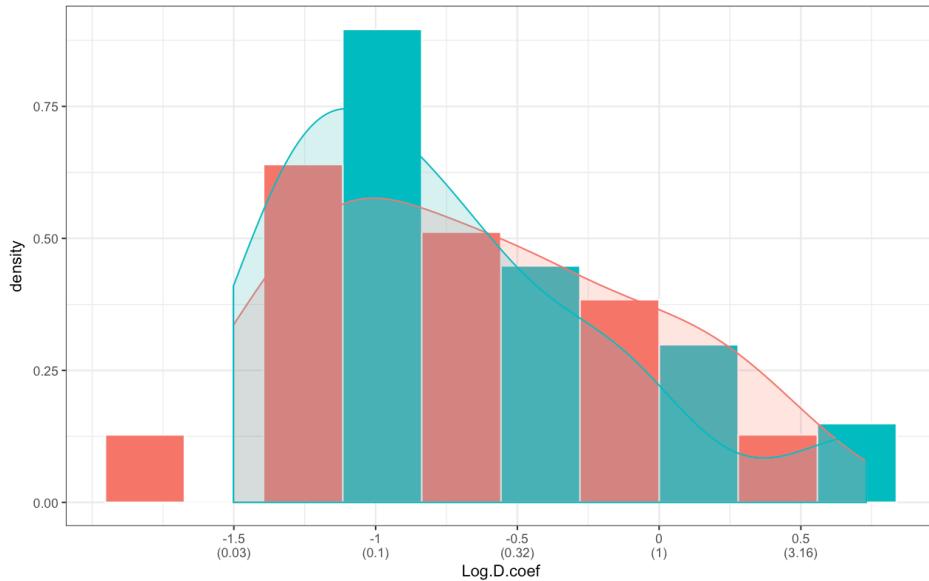
## Mean squared displacement

The diagram shows a particle's trajectory as a series of points labeled 1 through 8. The time interval between points is  $\Delta t$ . The displacement for each segment is calculated as the distance between consecutive points. The mean squared displacement (MSD) is then calculated as the average of the squared displacements for all segments.

$$\text{SD} = \sqrt{\frac{\sum d_i^2}{N}}$$
$$\text{MSD} = \frac{\sum_{i=1}^{n-t} d_{i,i+t}^2}{N}$$



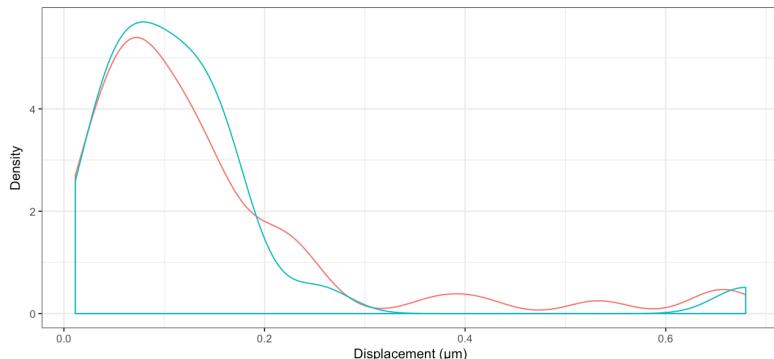
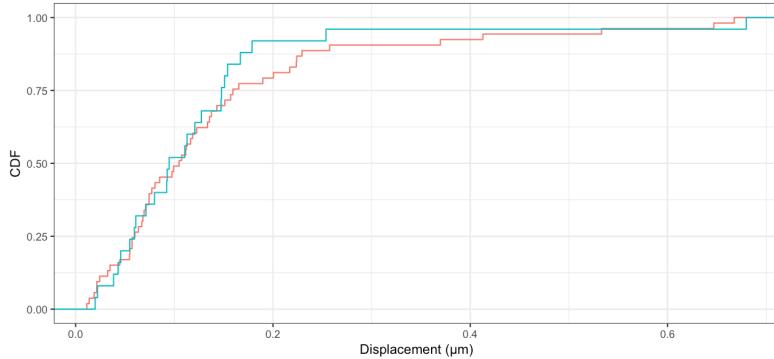
# Analyzing Data



***Diffusion coefficient***



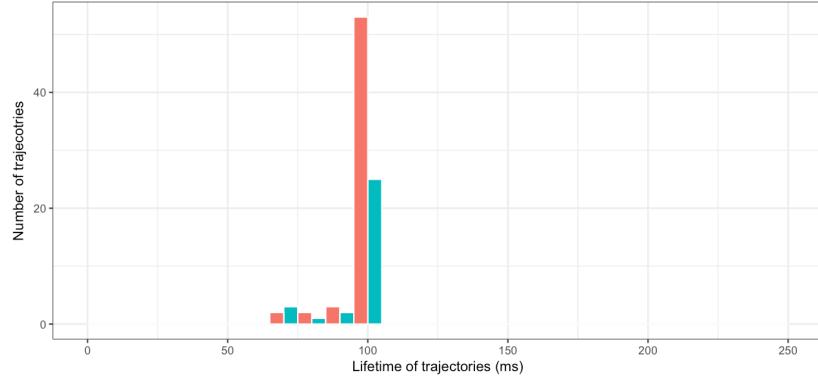
# Analyzing Data



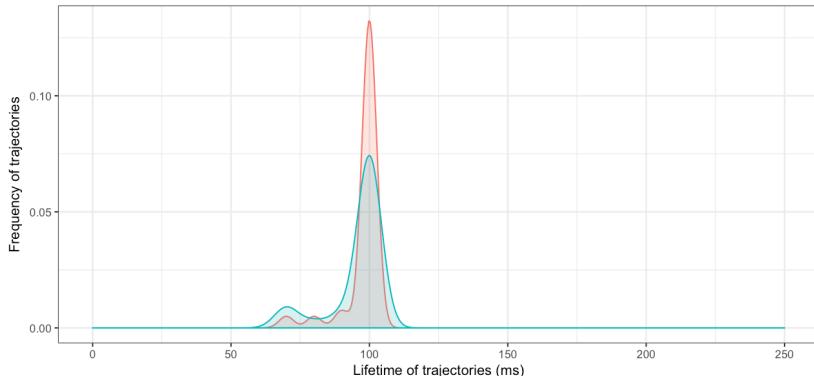
**Cumulative  
distribution function**



# Analyzing Data

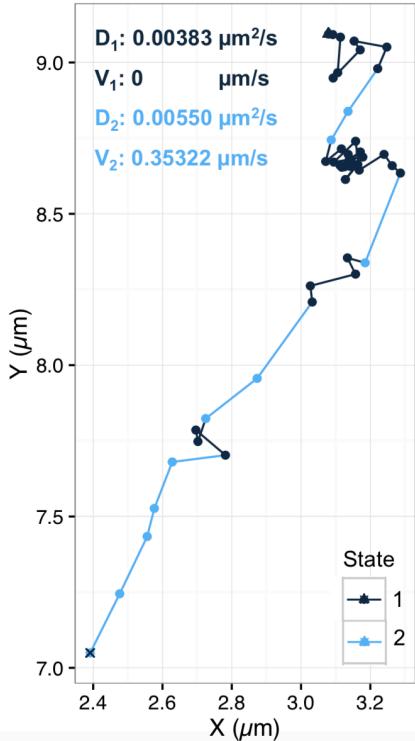


**Dwell time**





# Analyzing Data



**Hidden Markov  
models**

# Discussion

# Acknowledgements

## PI

**Prof. Carl Wu**

**Prof. TJ Ha**

**Prof. Michael Schatz**

## Post-Docs

**Sheng Liu**

**Anand Ranjan**

**Vu Nguyen**

**Jee Min Kim**

**Xiaona Tang**