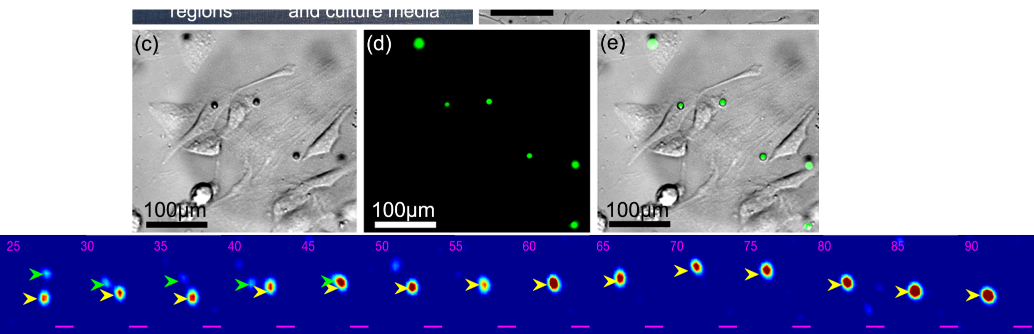
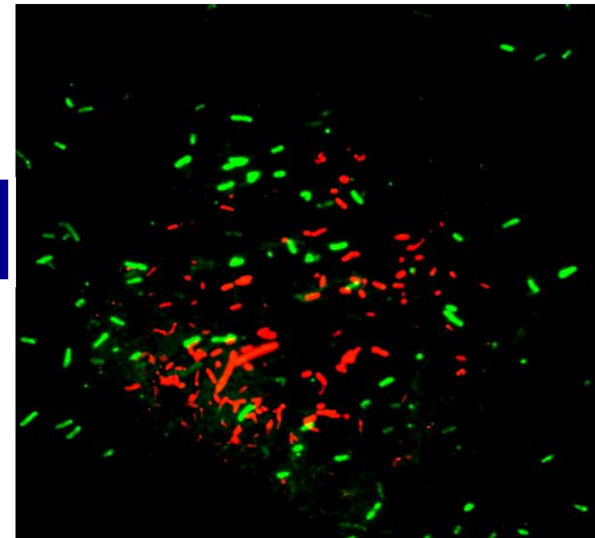
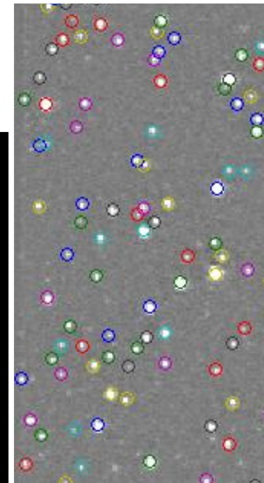
Lecture 11

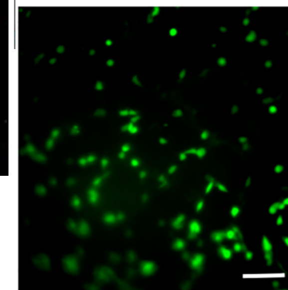
Image/video and data analysis

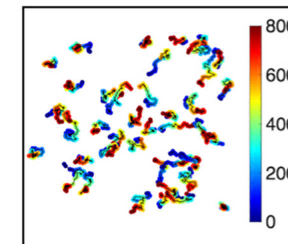
Halil Bayraktar
Moleculer Biology and Genetics



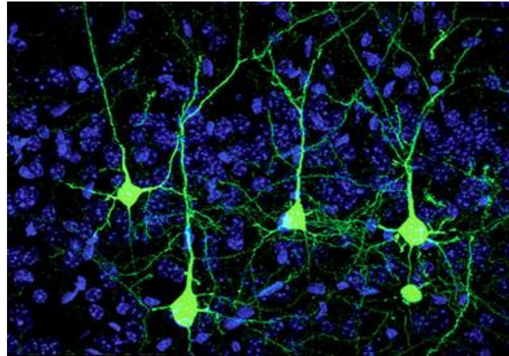


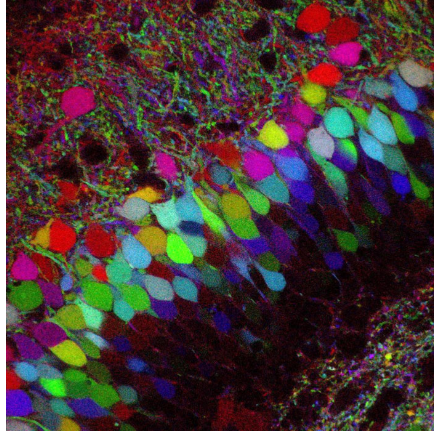




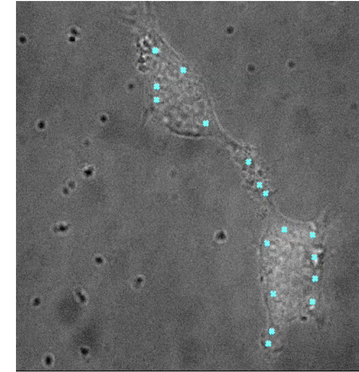


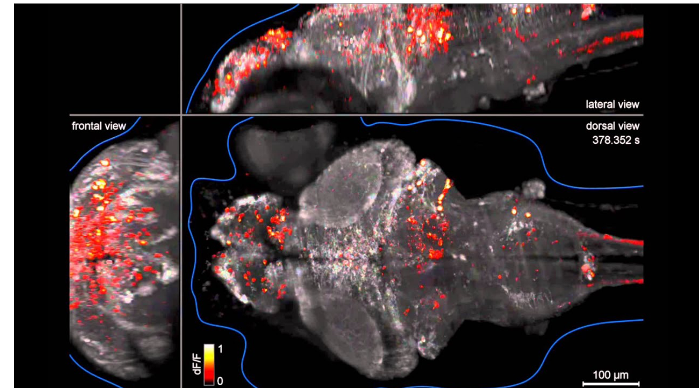
Optical tools and probes are needed to record signals at spatiotemporal resolution





Tamily A. Weissman[*,1] and Y. Albert Pan





Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M. and Keller, P. J. et al

GCaMP Fluorescent Reporters of Neuronal signaling

https://www.youtube.com/watch?v=FGvp6cdKb3c

Static Imaging of cell function
Low spatiotemporal resolution

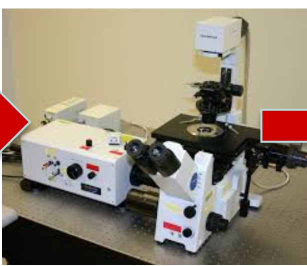
Dynamic Imaging of cellular function
High spatiotemporal resolution

Evolution of microscopes and technology for biological systems

1600 → 1960 → 2005 → 2012 → →

Zacharias Janssen

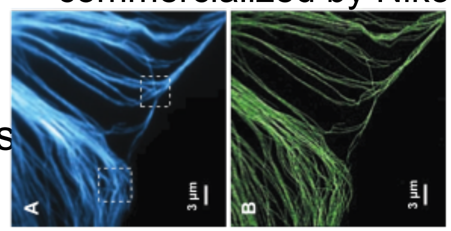
Upright Microscope

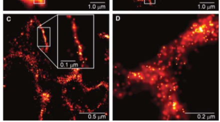


Marvin Minsky

Conventional confocal micros

Invented by Zhuang lab.
commercialized by Nikon Ins.

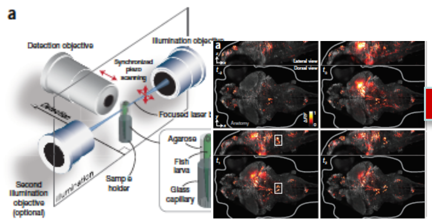




Invented by S. Hell Lab
Leica systems

Betzig lab

Super Resolution Microscope
(PALM, STED, STORM vs.)



Next?

Keller lab

Action potential
in zebrafish brain

Light Sheet Microscope
Commercially available

There is a strong interest for large field of view, functional imaging microscopy, organ on a chip based model imaging and advance computation.

As biologists demand new methods, novel tools will be invented.

Deep diagnostic with image processing

comment

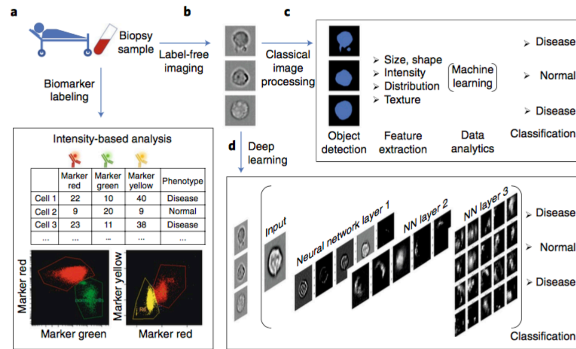


Fig. 1 | Valuable information is hidden in label-free images. High-throughput cell-based diagnostics allow samples to be analysed at single-cell resolution and across multiple channels. **a**, Conventional cell-based diagnostics often rely on specific biomarkers to identify disease status. The readouts are mainly intensity signals of the labelled targets. For multiplexed assays that involve several biological targets, such intensity-based analysis typically requires manual pairwise comparisons for the relevant markers. **b**, Recent research[9–11,13,14] indicates that label-free channels of images (such as brightfield and darkfield) can contain equivalent information, potentially replacing fluorescent markers. To accomplish this, however, it requires sophisticated extraction of information from images. **c**, In classical image processing pipelines, designed features (such as shape, intensity, texture) are helpful inputs for a machine classifier to learn the characteristic pattern of the phenotypes. However, feature engineering requires image analysis expertise and is limited in its maximum accuracy. **d**, In contrast, deep neural networks are generally more accurate and also more flexible: they identify features on their own by learning relevant patterns from a large number of examples (training dataset, not shown). One caveat is the loss of direct interpretability of the discovered features due to the hierarchy of abstract representation, as shown here on three hidden layers of a simple neural network.

hallmarks of disease. Screening cytology, such as Pap smears for abnormal cervical cells, enables life-saving early discovery of disease in the absence of clinical symptoms. As we will discuss, bringing machine learning to the analysis of microscopy/histology images offers tremendous potential for cell diagnostics with a greater ability to discern among patient subtypes.

For cells in suspension, such as in blood samples, flow-based systems are more favourable. Although flow cytometers are

at a throughput of several hundreds to thousands of objects per second. Signals from unwanted events, such as debris, can be more easily detected and ignored than in conventional flow cytometry. Currently, few cell-based diagnostics are in clinical use that rely on imaging flow cytometry, but as we will discuss, this is likely to change: the spatial information (that is, images) that imaging flow cytometry brings may soon reduce or eliminate the need for the specific biomarkers that are required for

This design can be highly modular and customizable, thus enabling parallelization and microcontrol of multiple functions in a single compact device, such as mixing, particle manipulation, imaging, tracking and other automated assays[2]. It can be used to study living cells together with their associated extracellular materials in the supernatant[3,4], which is much less feasible by microscopy or flow cytometry.

Trends in cell image analysis

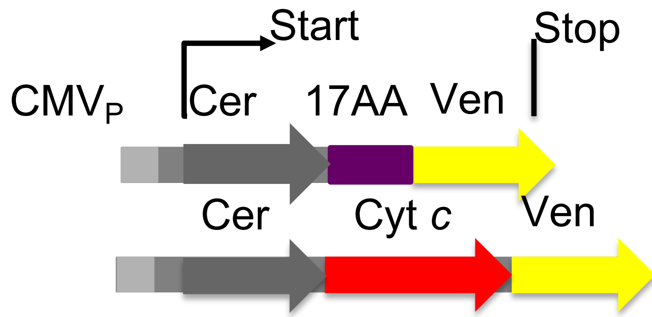
A dramatic revolution in computer vision has suddenly made new technology available for image analysis that, when combined with the image-capturing devices just described, could yield a crop of novel cell diagnostics.

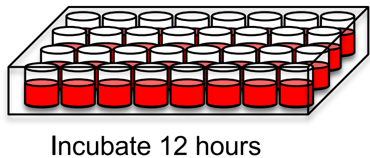
It is first helpful to understand existing approaches for analysing cell images for diagnostic purposes. Of course, the most widespread is the visual assessment of phenotypes by pathologists. This raises challenges: trained experts are expensive and cannot analyse enormous datasets efficiently, as in whole-slide scans of a tissue biopsy, for example. Furthermore, discrepancies among pathologists' judgment are well-documented[5], and it is possible that patterns exist in cell morphology that the human visual system is simply not equipped to perceive[6].

Image analysis software can overcome many of these challenges. In classical image processing (Fig. 1c) a researcher designs algorithms to identify each cell, its borders and any relevant subcellular compartments (for example, nuclei or other organelles) in the images so that many different kinds of measurements of these identified regions of the image can then be taken. These so-called morphological features include pixel intensities, size, shapes, textures, correlations and relationships among neighbour cells and subcellular components; these can be used directly as a diagnostic feature. Features can also be combined to detect more complex phenotypes that manifest in multiple features simultaneously using machine learning, where the algorithm learns to

Nature, Carpenter et. al.

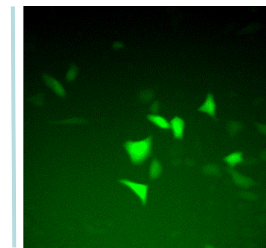
Data volume: high-content screening studies generate bid data sets





DNA+ Transfection Reagent

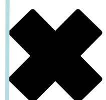
Fluorescence microscopy



78 frames

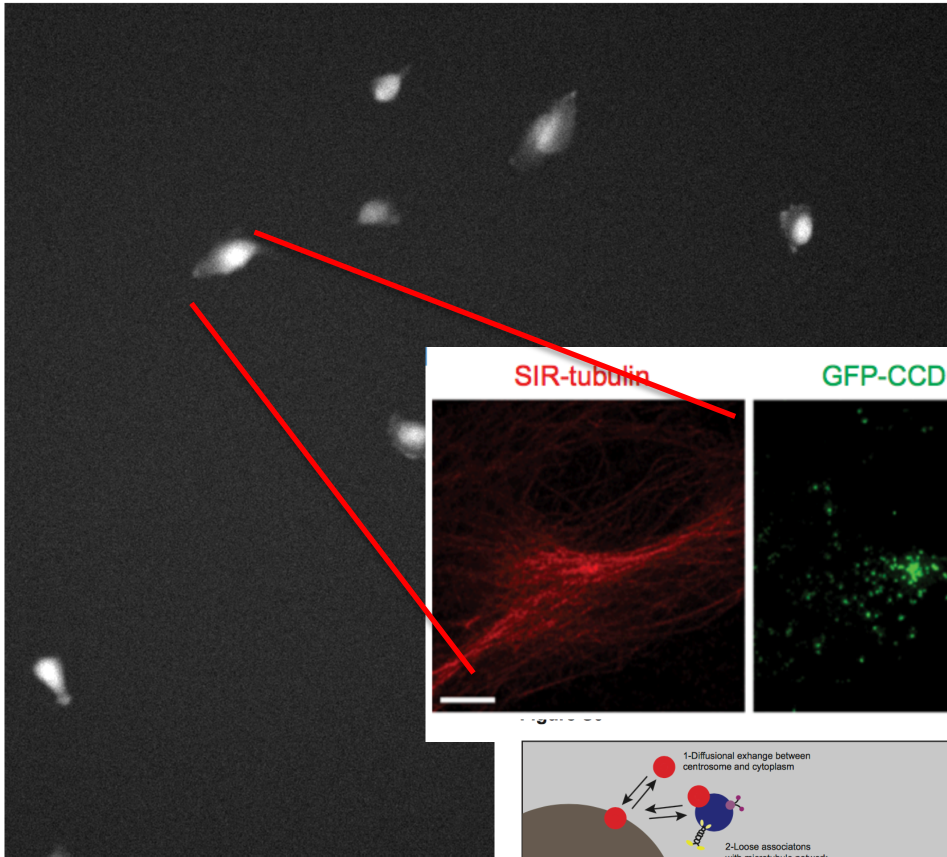
FRET Samples (include controls)

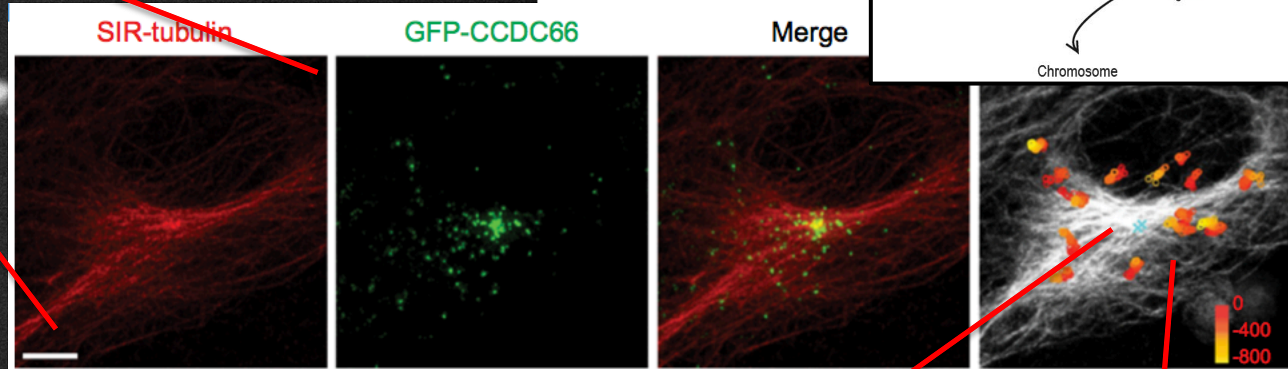
180 spots

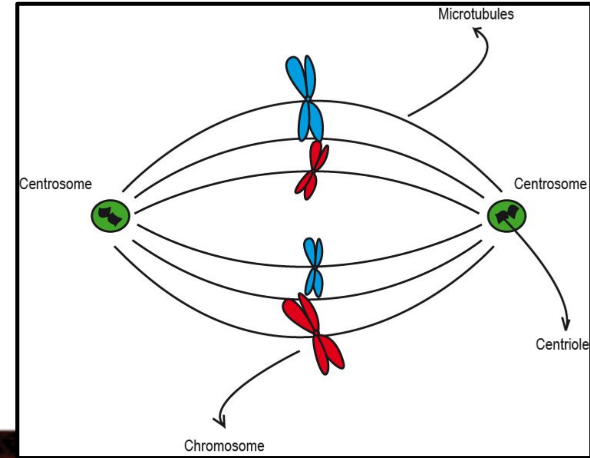
 4 channels

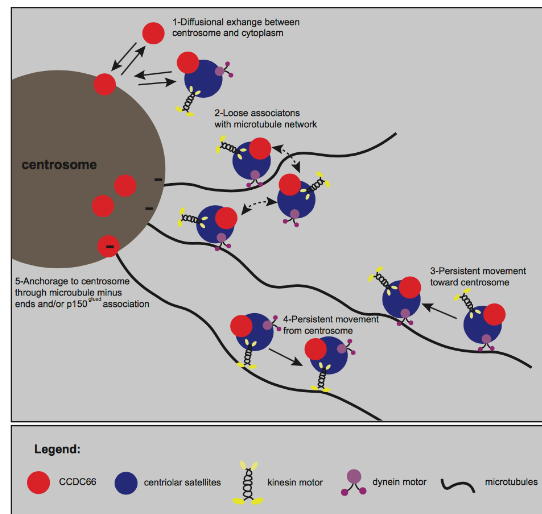
= ~ 1.5×10^{10} data points / day

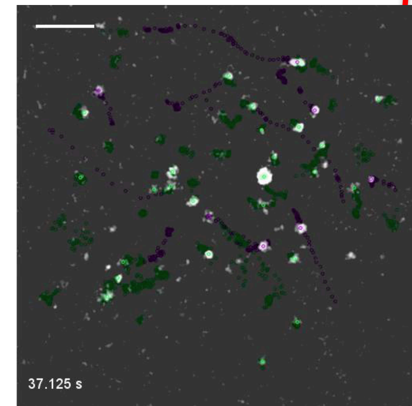
Imaging cells at different scales





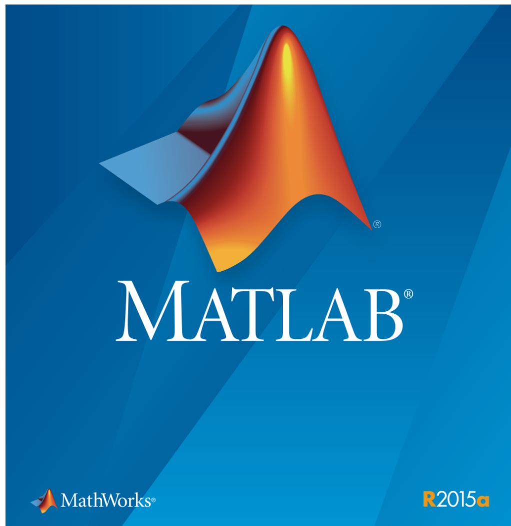






Conkar et al. scientific reports 2019

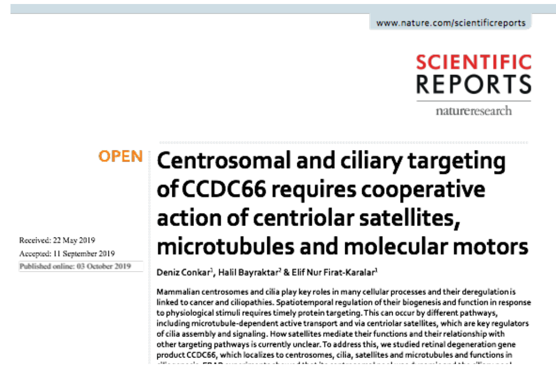
Custom programs for analysis of biological systems



Where do we use custom programming to determine the properties/dynamics of biological systems

1. Image/video processing
2. Genome analysis
3. Microarray analysis
4. Proteomics analysis
5. Advance graphics

Example: Understanding satellite dynamics leads us to develop new computer algorithms.


www.nature.com/scientificreports

SCIENTIFIC REPORTS

natureresearch

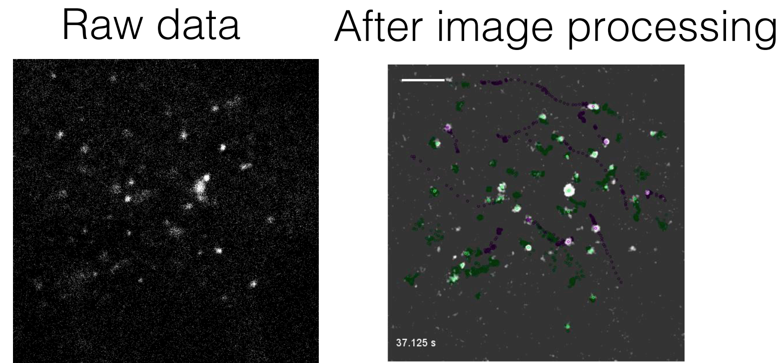
OPEN

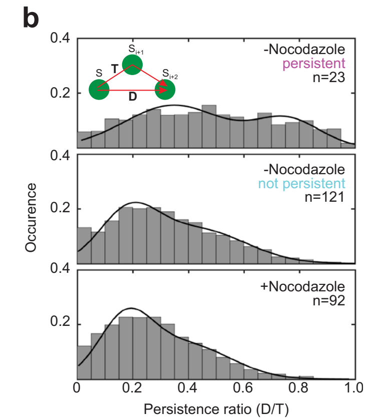
Received: 22 May 2019
Accepted: 11 September 2019
Published online: 03 October 2019

Centrosomal and ciliary targeting of CCDC66 requires cooperative action of centriolar satellites, microtubules and molecular motors

Deniz Conkar[1], Halil Bayraktar[2] & Elif Nur Firat-Karalar[1]

Mammalian centrosomes and cilia play key roles in many cellular processes and their deregulation is linked to cancer and ciliopathies. Spatiotemporal regulation of their biogenesis and function in response to physiological stimuli requires timely protein targeting. This can occur by different pathways, including microtubule-dependent active transport and via centriolar satellites, which are key regulators of cilia assembly and signaling. How satellites mediate their functions and their relationship with other targeting pathways is currently unclear. To address this, we studied retinal degeneration gene product CCDC66, which localizes to centrosomes, cilia, satellites and microtubules and functions in





Challenges on the project:

1. Noisy images/videos
2. Moving objects
3. New parameters are needed to determine satellite dynamics
4. Advance graphics for data visualization

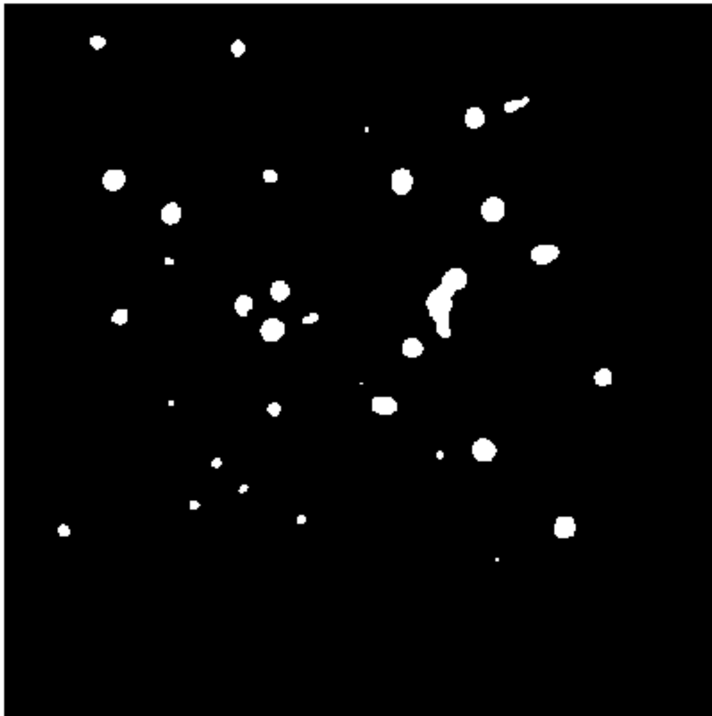
Our solutions was;

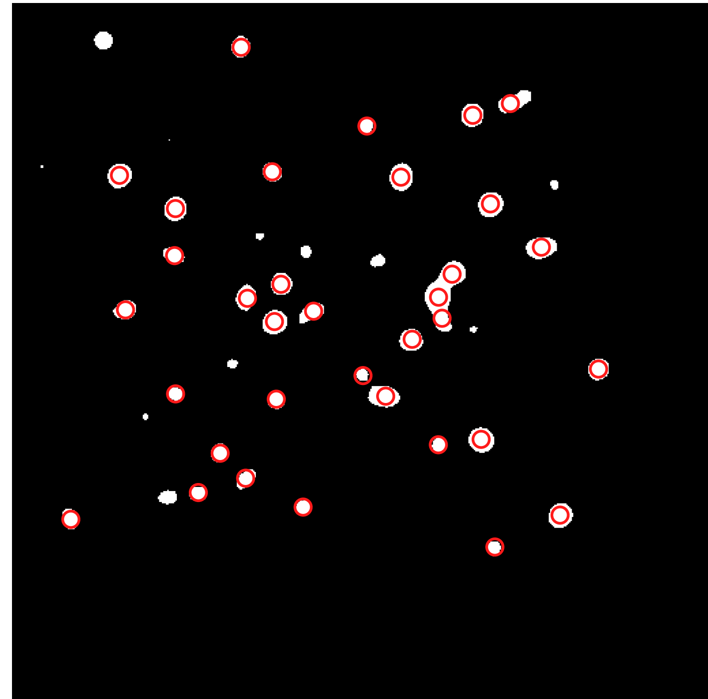
1. Noise free images
2. Build a custom tracking algorithms for moving objects (Satellites)
3. Compute new parameters : persistence, speed, distance, number etc.
4. Custom solutions for data visualization

Result: Analysis demonstrated that satellites can be distinguished based on their persistence ration and around centrosome they move both diffusively and persistently

Finding objects in images: Threshold filter

A cutoff intensity filter is used to determine the locations of satellites in images





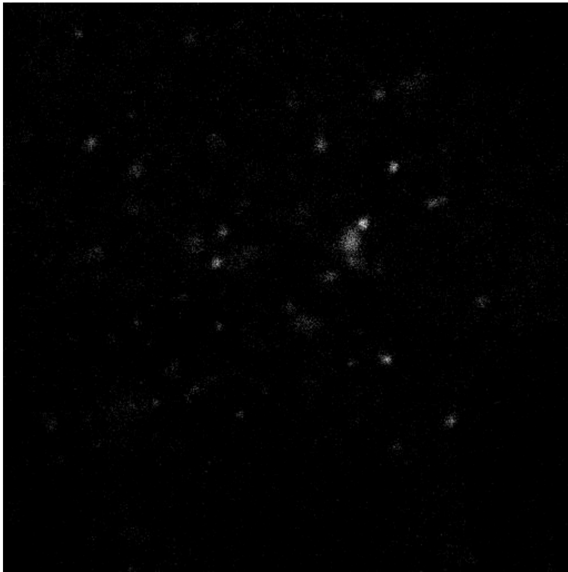
Challenges in Fluorescence Microscopy

Although we use high-end and super expensive microscopes, they are not perfect.

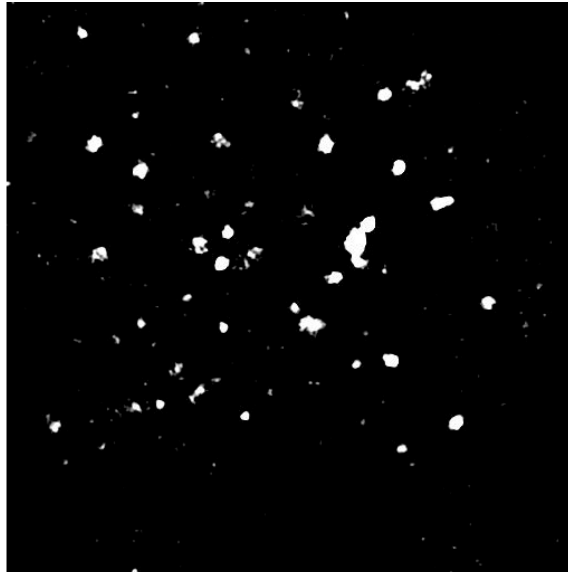
- Low signal to noise ratio
- Some issues: Blur images, pixel noise, focus loss, diffraction issues etc.

- Solution: Post-processing of image/videos

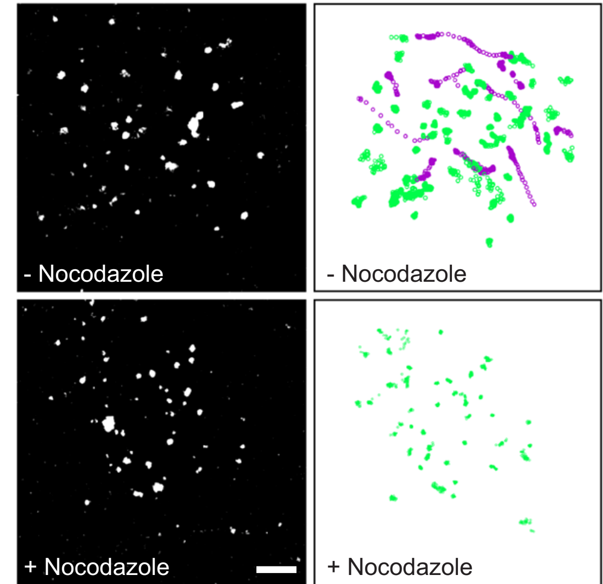
Technical term: Bandpass (low-pass) filter was used to remove noise.



Raw data:Pre-processing

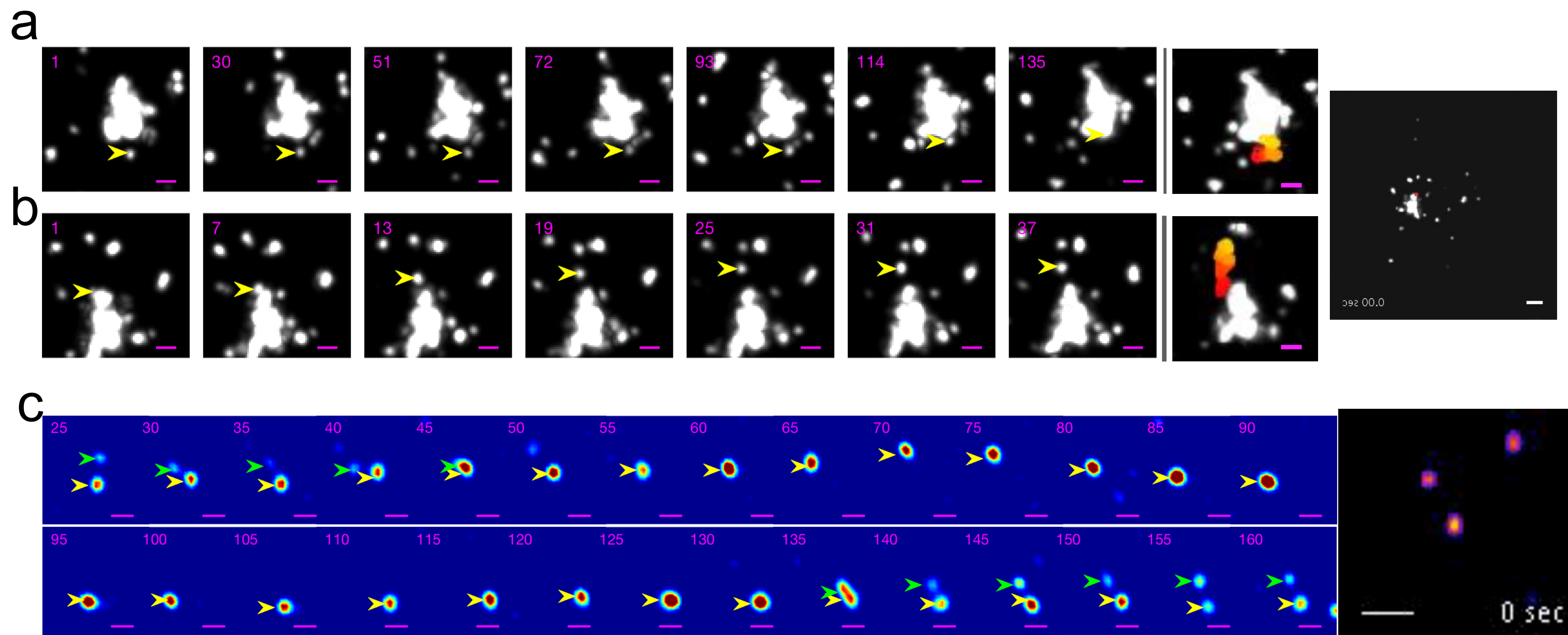


Post-processing



Conkar et al. scientific reports 2019

Graphics for better data presentation



d

Fusion Splitting

Fluorescence intensity (a.u.)

1
0.8
0.6
0.4
0.2
0

0 45 90 135 180

Time (s)

Conkar et al. 2019

Static and dynamic information with high spatial and temporal resolution

Dynamic at time proint t
-Average velocity
-Instant speed
-Persistence
-Direct distance
-Total distance

Other dynamic properties
-Size
-Shape
-Gene expression levels

How these properties/features change as a function of different perturbation?
How much they different across different cell types?
How do they affect cell fate?
Are they different in normal and cancerous cells?

Some available function for image analysis

imread	Read an image in a variety of formats
imfinfo	Gather information about an image file
imwrite	Write data to an image file
image	Display image from array
imshow	Display an image, optimizing figure, axes, and image object prop- erties, and taking an array or a filename as an input
rgb2gray	Rgb to gray scale

Reading tif file name

- Spfile=dir('*.tif')

5x1 struct with 6 fields

Fields	name	folder	date	bytes	isdir	datenum
1	'Mark_and_Find 001_Position007_t000_R...	'/Users/...	'23-Jan-...	5532654	0	7.3781e+05
2	'cellimage.tif'	'/Users/...	'15-May...	5322953	0	7.3793e+05
3	'example1.tif'	'/Users/...	'13-May...	210949	0	7.3792e+05
4	'example2.tif'	'/Users/...	'13-May...	185639	0	7.3792e+05
5	'example3.tif'	'/Users/...	'13-May...	193645	0	7.3792e+05
6						
7						
8						

Each image comes with a metadata that demonstrates camera software, image properteis, where and how the image was generated.

This is useful when analyzing images and videos

```
Spfile(1).name
a =
imfinfo(Spfile(1).name)
```

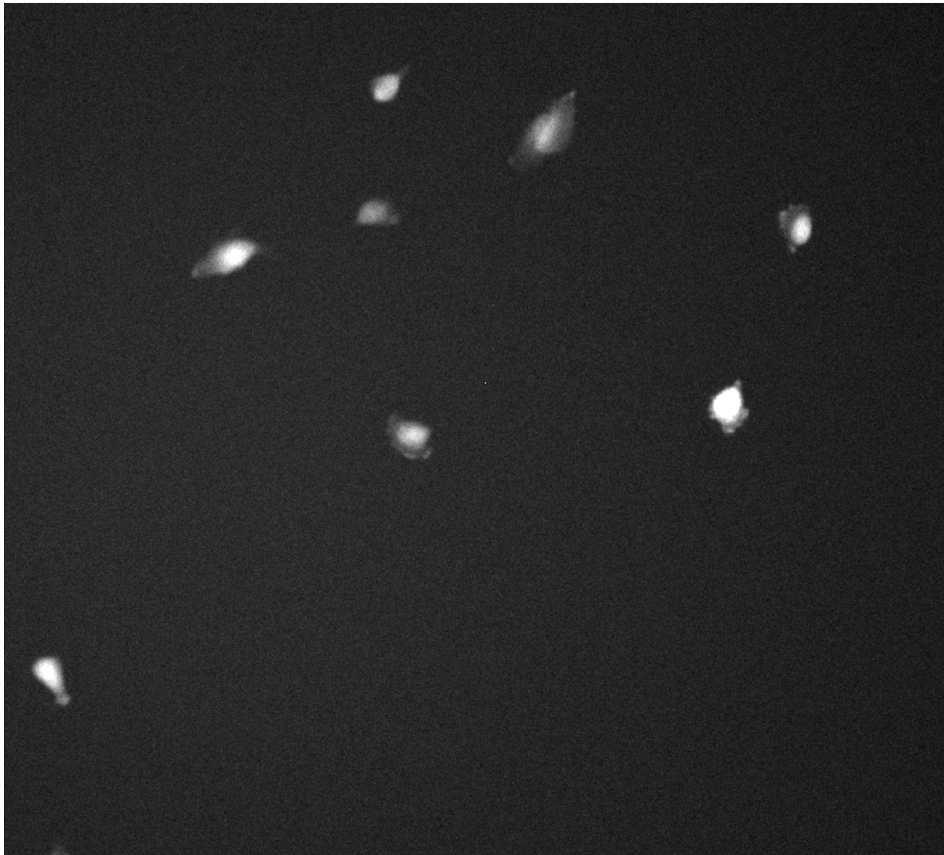
Field	Value
Filename	'/Users/halilbayraktar/Documents/Teaching/Scientific Computation...
FileModDate	'23-Jan-2020 16:04:24'
FileSize	5532654
Format	'tif'
FormatVersion	[ ]
Width	1920
Height	1440
BitDepth	16
ColorType	'grayscale'
FormatSignature	[73,73,42,0]
ByteOrder	'little-endian'
NewSubFileType	0
BitsPerSample	16
Compression	'Uncompressed'
PhotometricInte...	'BlackIsZero'
StripOffsets	1x360 double
SamplesPerPixel	1
RowsPerStrip	4
StripByteCounts	1x360 double
XResolution	1.5422e+04
YResolution	1.5422e+04
ResolutionUnit	'Centimeter'
Colormap	[ ]
PlanarConfigura...	'Chunky'
TileWidth	[ ]
TileLength	[ ]
TileOffsets	[ ]
TileByteCounts	[ ]
Orientation	1
FillOrder	1
GrayResponseUnit	0.0100
MaxSampleValue	65535
MinSampleValue	0
Thresholding	1

Size of an image

```
xsize = a(1).Width;
ysize = a(1).Height;
```

1920 pixels

1440 pixels

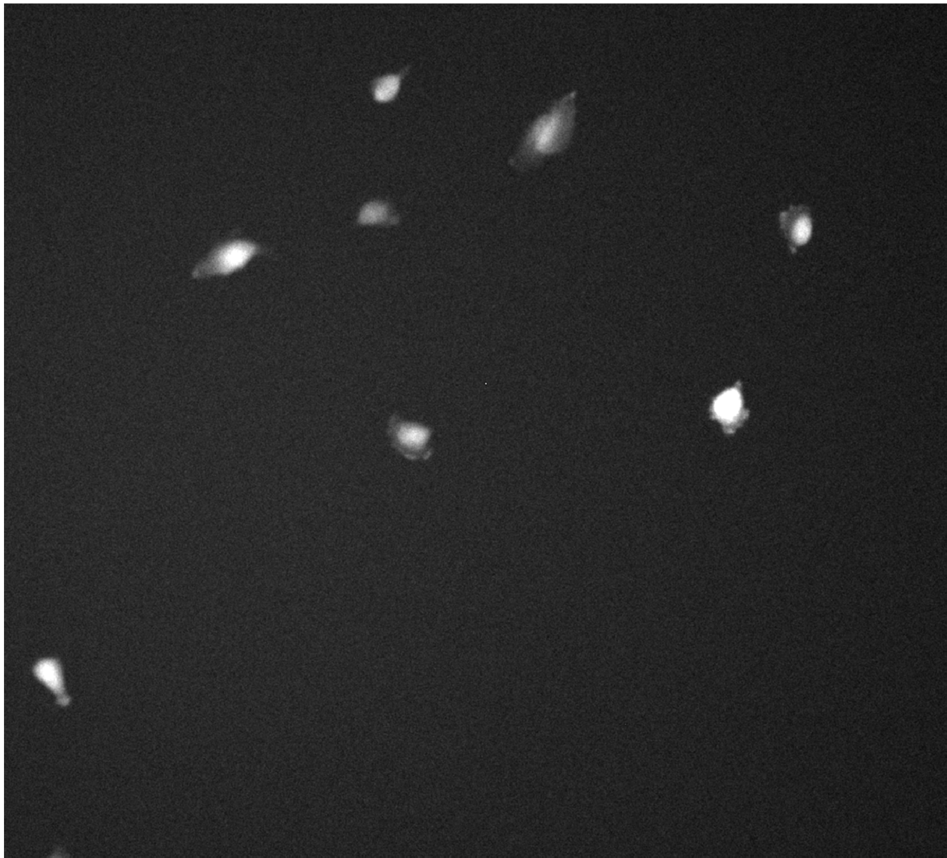


Field	Value
Filename	'/Users/halilbayraktar/Documents/Teaching/Scientific Computation...
FileModDate	'23-Jan-2020 16:04:24'
FileSize	5532654
Format	'tif'
FormatVersion	[ ]
Width	1920
Height	1440
BitDepth	16
ColorType	'grayscale'
FormatSignature	[73,73,42,0]
ByteOrder	'little-endian'
NewSubFileType	0
BitsPerSample	16
Compression	'Uncompressed'
PhotometricInte...	'BlackIsZero'
StripOffsets	1x360 double
SamplesPerPixel	1
RowsPerStrip	4
StripByteCounts	1x360 double
XResolution	1.5422e+04
YResolution	1.5422e+04
ResolutionUnit	'Centimeter'
Colormap	[ ]
PlanarConfigura...	'Chunky'
TileWidth	[ ]
TileLength	[ ]
TileOffsets	[ ]
TileByteCounts	[ ]
Orientation	1
FillOrder	1
GrayResponseUnit	0.0100
MaxSampleValue	65535
MinSampleValue	0
Thresholding	1

Read images and show it in the figure

```
datB = imread(Spfile(1).name,
'tif', 1);
```

```
figure(1)
ax=imshow(datB,[min(min(datB)) max(max(datB))/3])
```



Black-and-White Images

It is a 1 bit image. The pixel can carry either 0 (black) or 1 (white).

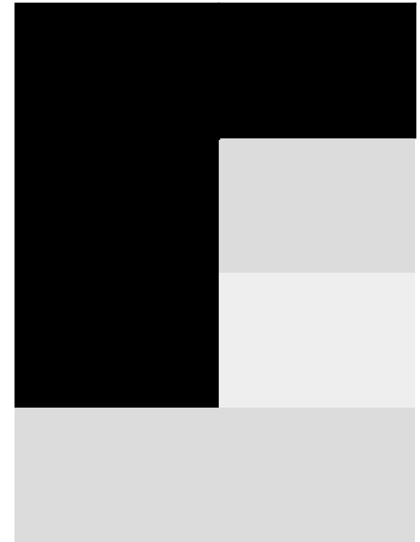
That is also called binary image.

No gray levels from black to white is present.

Binary Image

RGB Image

0	**1**	**0**	**0**
0	1	0	1
0	1	0	1
0	1	1	1



Gray Scale Images:

It is an image where the intensity values are scaled between black and white.

If only two color is available for pixels. It is a 1-bit picture- Pixels are 0 or 1
If 256 colors it is 8 bit (1byte) picture from
0 to 255. 8 bit or 1 Byte image has a 256 shades of gray if a gray scale image used.

00000000 = 0 black
.
00111110 = 62 gray tone
.
.
11111111 = 255 white

You can represent any number from 0 to 255 using 8 bits.

6000x4000 pixels

Numbers represents the 256 shades of gray intensity values from 0 up to 255.

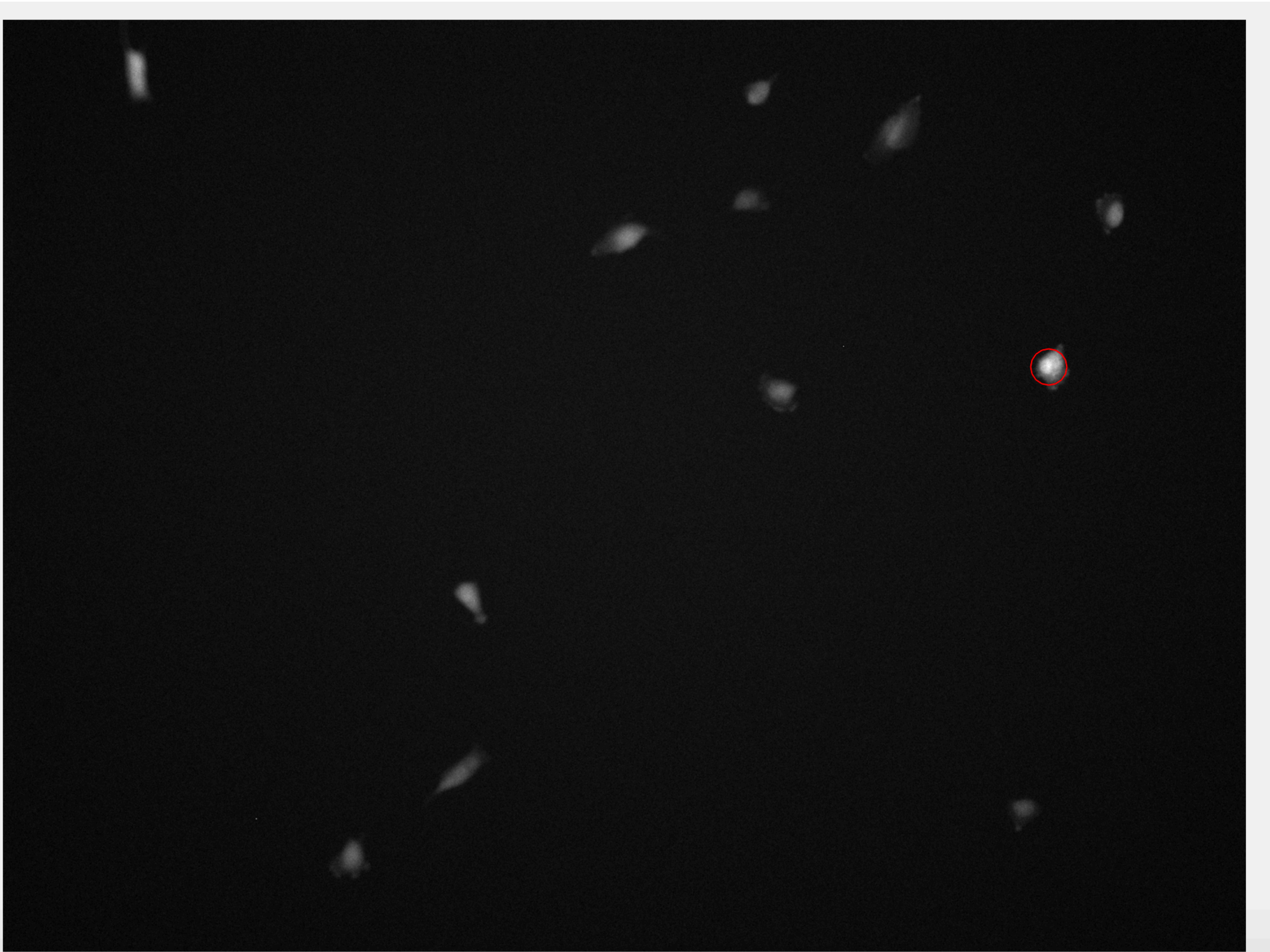
1 byte = 8 bit image

2^8 = 256 different levels

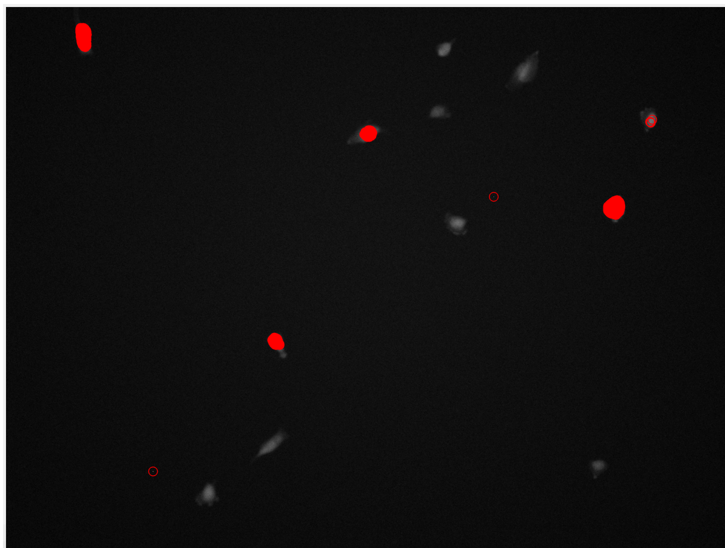
10 by 10 pixel – upper corner of the image

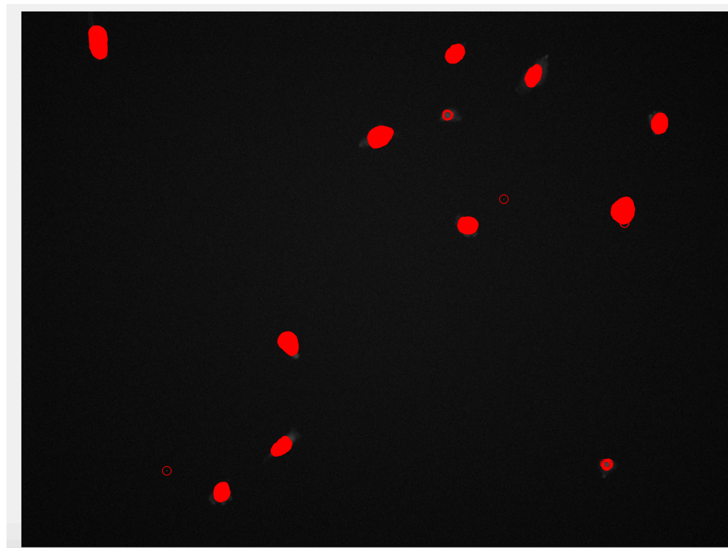
85	78	75	79	82	81	80	81	81	89
74	65	57	56	55	52	49	50	71	74
73	63	56	56	58	57	58	61	66	62
64	56	52	55	59	61	63	66	61	54
68	60	54	56	57	55	53	54	53	47
73	64	57	56	57	54	53	54	51	49
66	55	45	43	45	47	50	54	51	54
80	66	52	47	47	49	54	60	47	55
83	73	59	49	48	53	58	60	55	52
74	67	61	58	53	48	49	54	54	52

Finding brightest cell in the image



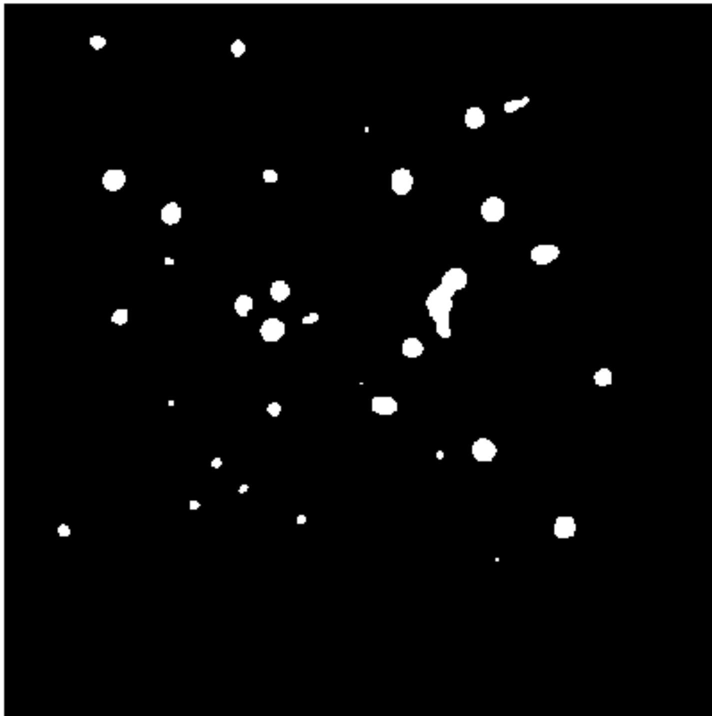
Finding all cells in the image

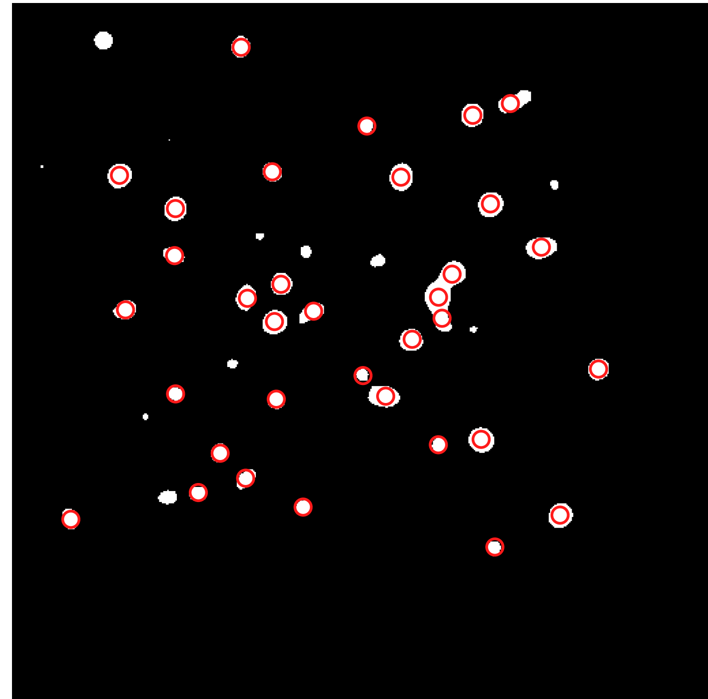




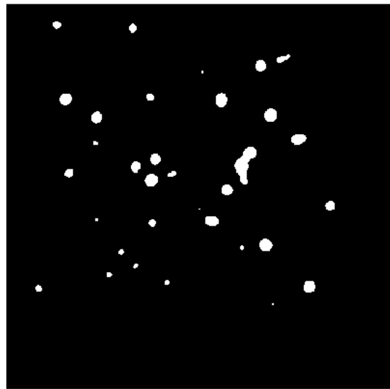
Finding objects in images: Threshold filter

A cutoff intensity filter is used to determine the locations of satellites in images

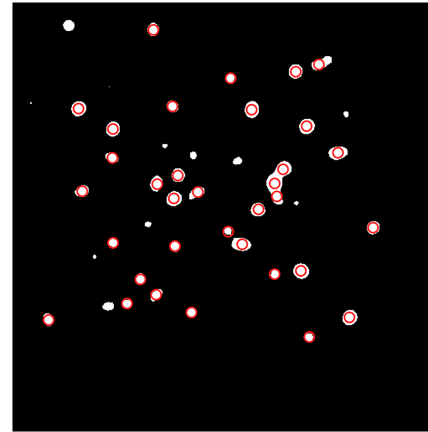




How many spots/protein complex are present?



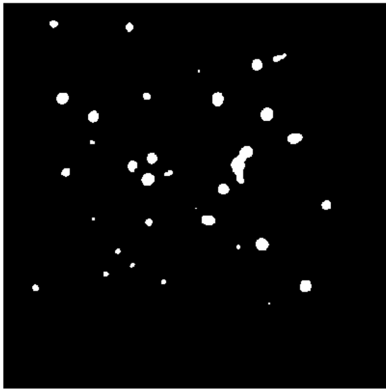
Threshold filter



Intensity > threshold

Size filter (pixels)

1	67	244.552	231	255	56.642	88.736
2	81	246.444	232	255	133.458	99.893
3	109	245.514	230	255	188.591	105.818
4	132	246.311	231	255	82.823	141.115
5	38	245.289	231	255	187.224	134.800



- -Size (5 to Infinity)
- -Show outline
- -Set measurements

Many information can be extracted

Area = Number of Pixels

Diameter = Number of Pixels

	Area	Mean	Min	Max	XM	YM	Diameter	Perimeter
1	67	244.552	231	255	56.642	88.736	12	33
2	81	246.444	232	255	133.458	99.893	13.9	38
3	109	245.514	230	255	188.591	105.818	18	47
4	132	246.311	231	255	82.823	141.115	25	60
5	38	245.289	231	255	187.224	134.800	9	22

Why we transform images?

To extract and interpret the information in the data

Many protein complexes
What information we can get from these segmentation?

1. Size
2. Intensity
3. Shape
4. Diameter
5. Total number in different cell lines
6. Extract dynamic information