Intro to Measurements & Quantification of Images

What information is in images and what isn't?

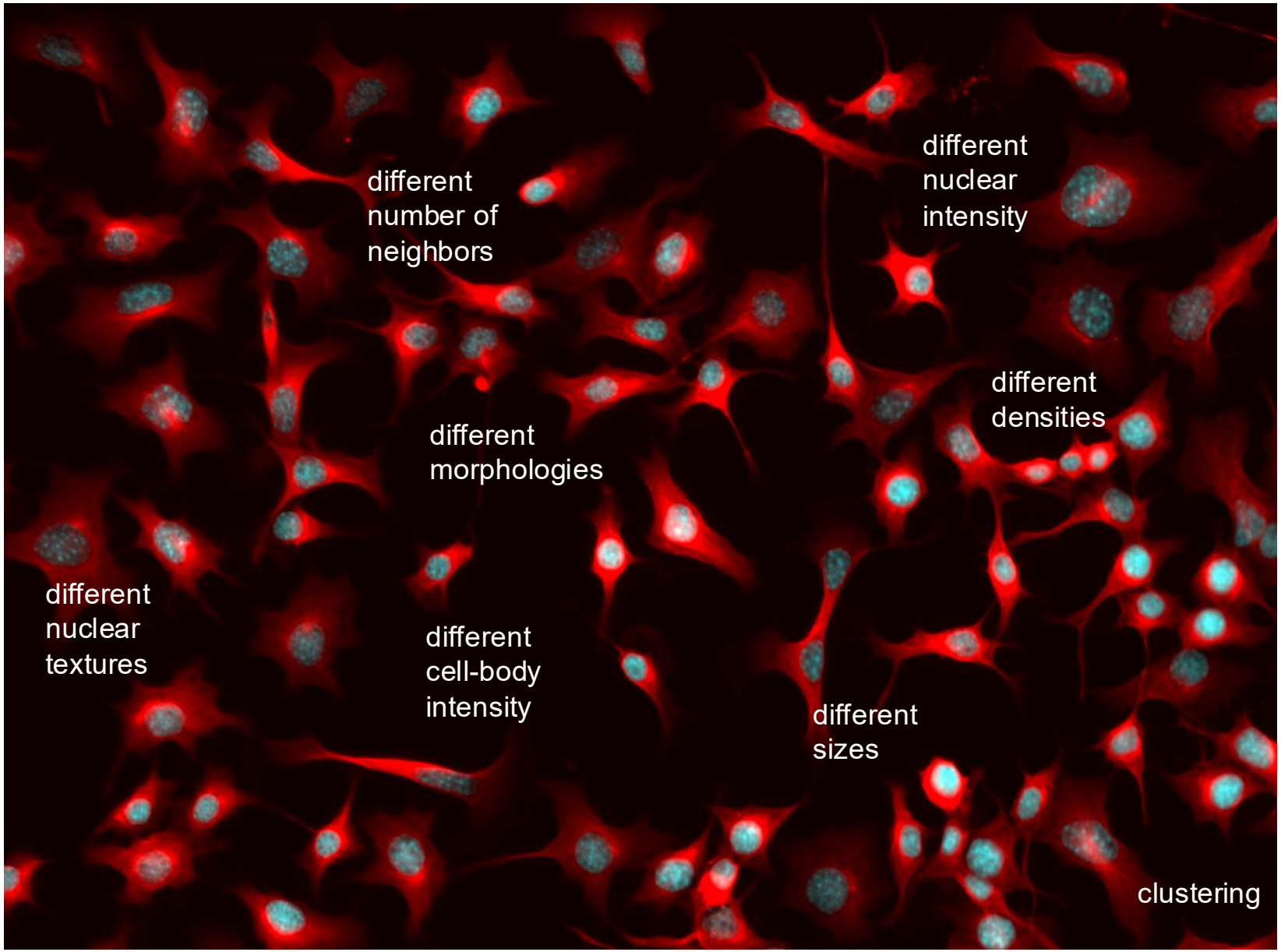
MAX BRAMBACH | OYLER-YANIV LAB

What do you see in this image?

When we describe images we tend to implicitly describe measurements on images.

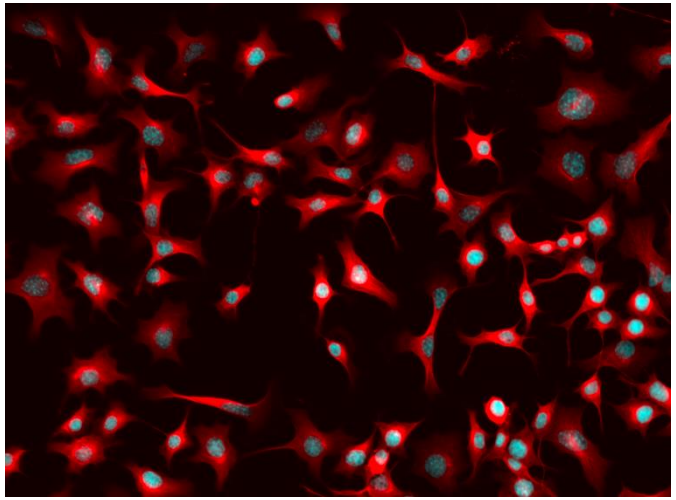
To draw conclusions from images we need to standardize how we describe images.

We measure properties and make our observations solid and explicit.



Images are not data

What we see



Image

What we actually have

```
array([[ 98, 105, 107, ..., 122, 122, 122],
       [119, 120, 118, ..., 136, 134, 139],
       [135, 134, 133, ..., 143, 141, 143],
       ...,
       [164, 161, 161, ..., 383, 373, 354],
       [159, 160, 160, ..., 361, 363, 359],
       [160, 157, 161, ..., 360, 355, 361]],
      shape=(1040, 1392), dtype=uint16)
```

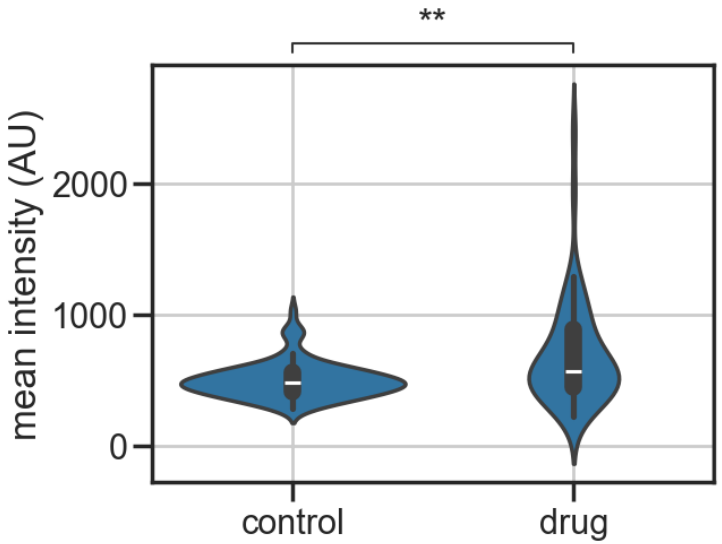
Array of intensity values

What we need

	label	area	mean_dapi	mean_protein
0	1	889.0	0.438285	0.253473
1	2	1245.0	0.357890	0.331311
2	3	609.0	0.448923	0.352874
3	4	1005.0	0.406975	0.341602
4	5	697.0	0.360550	0.341725
5	6	697.0	0.445584	0.264794

Measurements

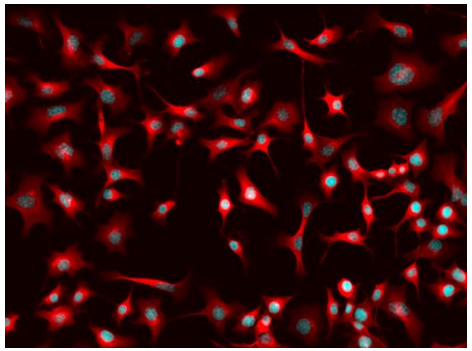
What we want



Results

From images to data is a series of choices

Image



Your scientific question

What **to measure**

- field of view
- individual cells
- cell groups
- subcellular objects
- punctae

= the objects

What to *measure*

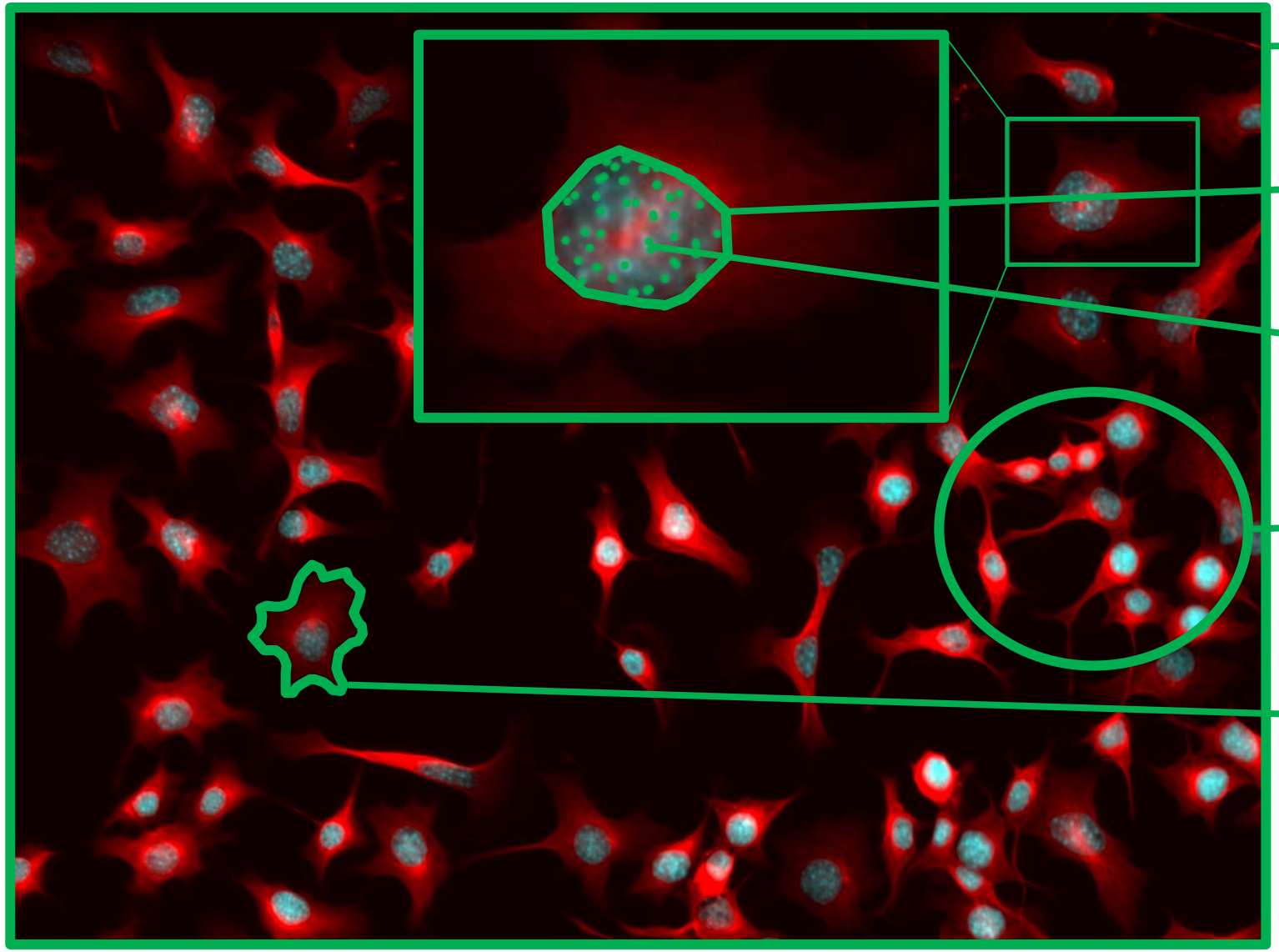
- number
- fluorescence intensity
- morphology
- density
- co-localisation

= the properties

Data

	label	area	mean_dapi	mean_protein
0	1	889.0	0.438285	0.253473
1	2	1245.0	0.357890	0.331311
2	3	609.0	0.448923	0.352874
3	4	1005.0	0.406975	0.341602
4	5	697.0	0.360550	0.341725
5	6	697.0	0.445584	0.264794

What could be an object?



Whole field of view

Nuclei

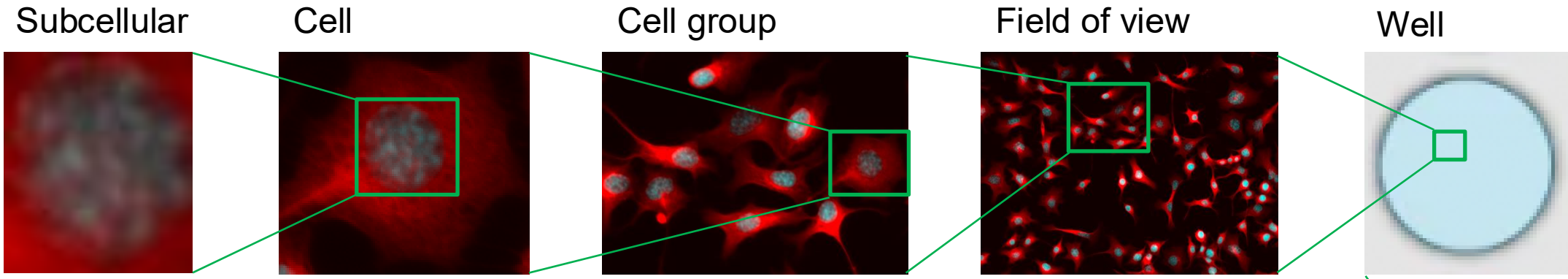
Puncta /
subcellular structures

Cell groups / clusters

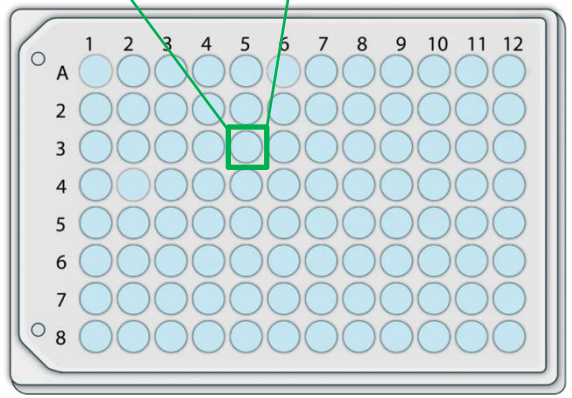
Single cells

Which objects to measure is a choice

Objects are hierarchical / nested



Plate/experiment



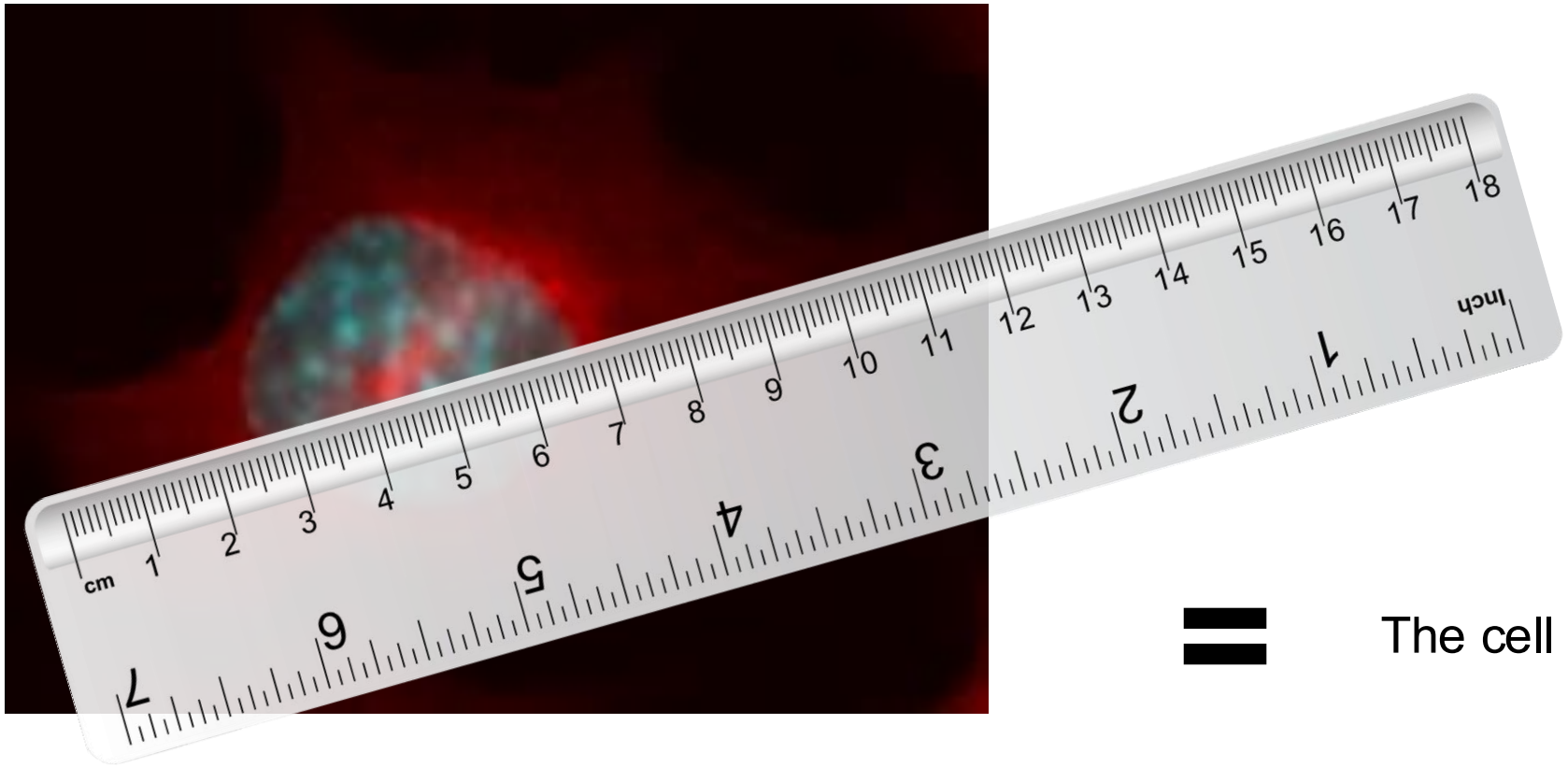
The choice of objects determines what can be asked.

- How many puncta per cell?
- How many cells per cluster?
- How many clusters per well?

What determines which object to pick?

YOUR SCIENTIFIC QUESTION

A measurement extracts a meaningful number from an object

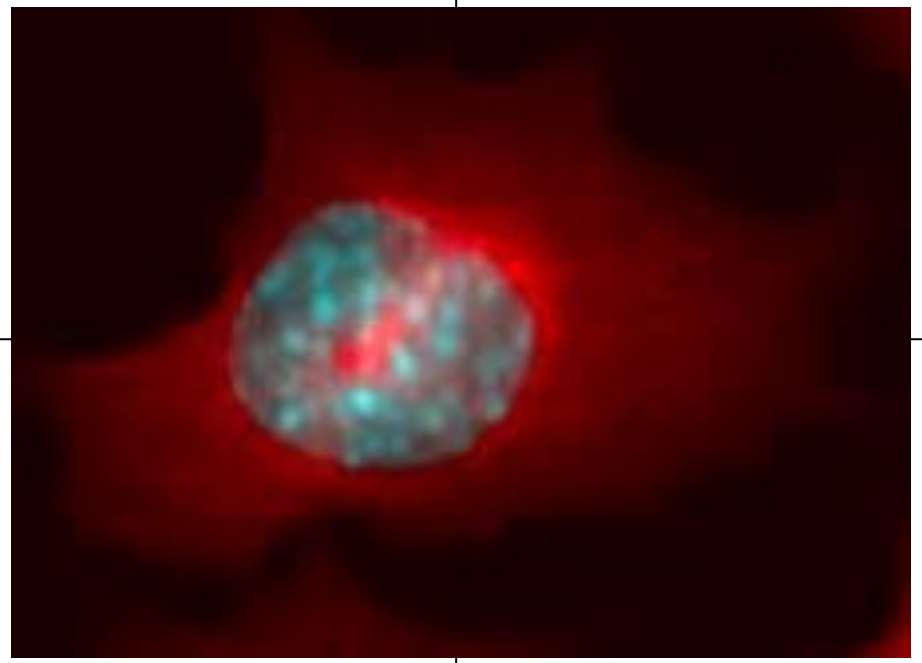


= The cell has a diameter of 11 µm.

There are four common categories of direct measurements

Count

e.g., how many dots?



Intensity

e.g., how bright is the staining?

Location

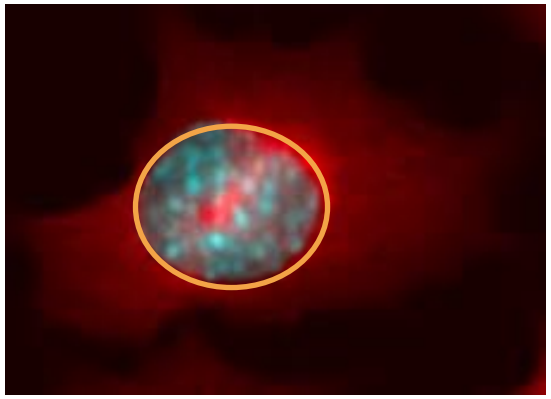
e.g., what are the coordinates of the dots?

Shape

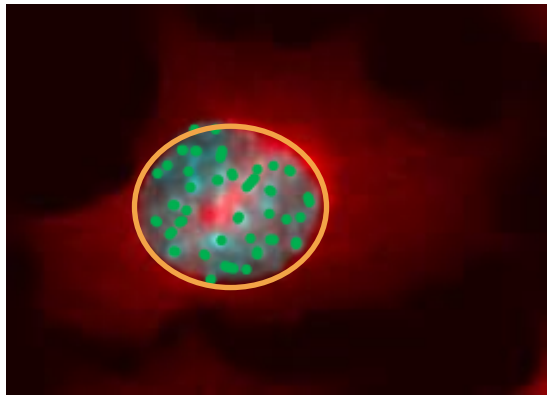
e.g., what is the size of a cell?

Measurements we care about often need to be calculated

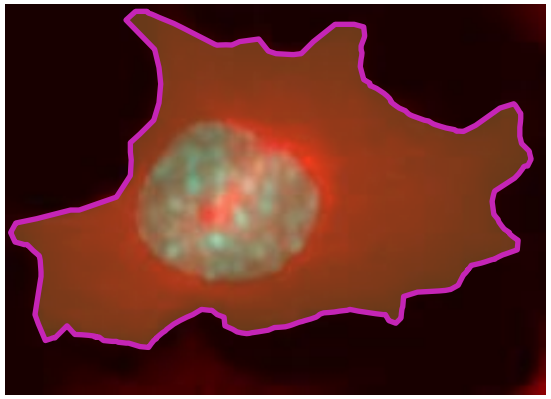
Intensity / Area = average intensity



Count / Area = density



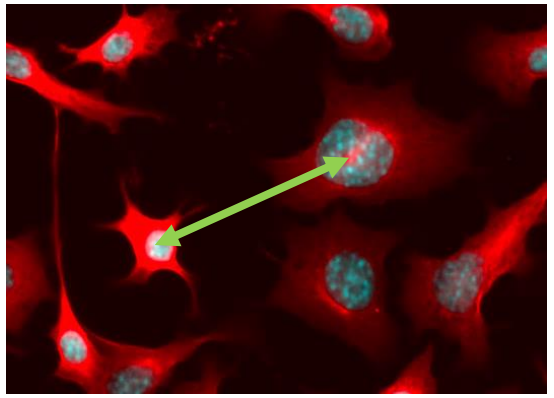
Perimeter^2 / Area = morphology description



circle: 4π

square: 16

location → location = distance

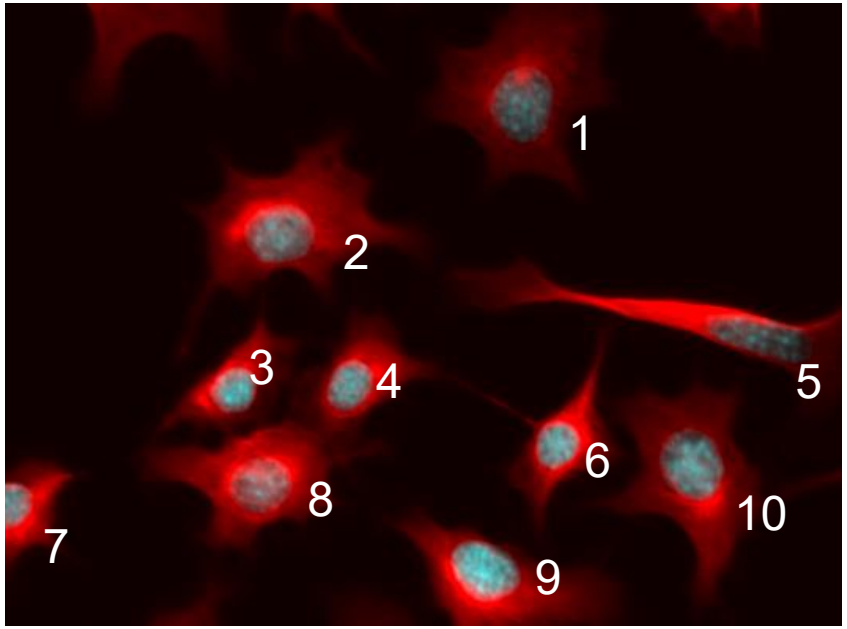


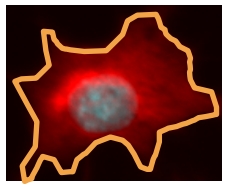
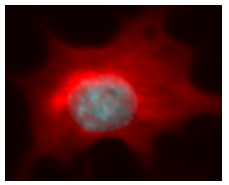
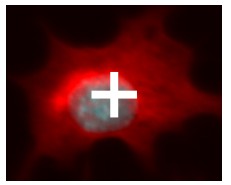
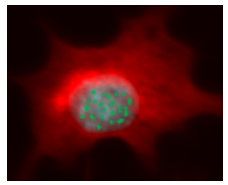
What to measure is a choice.

What determines which measurements to pick?

YOUR SCIENTIFIC QUESTION

Data are tables



MEASUREMENTS

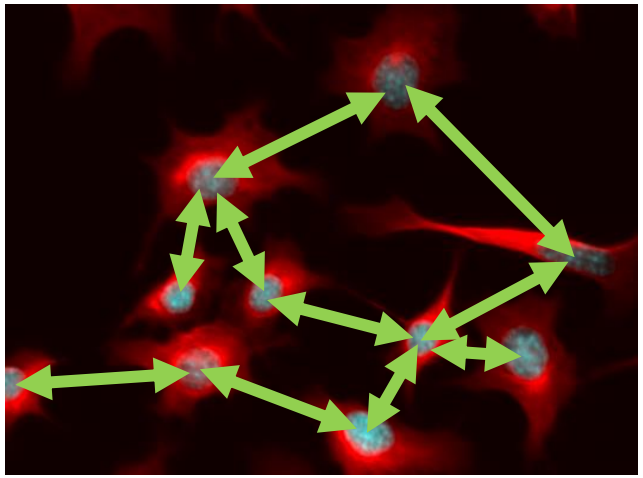
OBJECTS

object id	area µm²	intensity ch. 1	intensity ch. 2	coord. x	coord. y	No. of spots
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

DATA

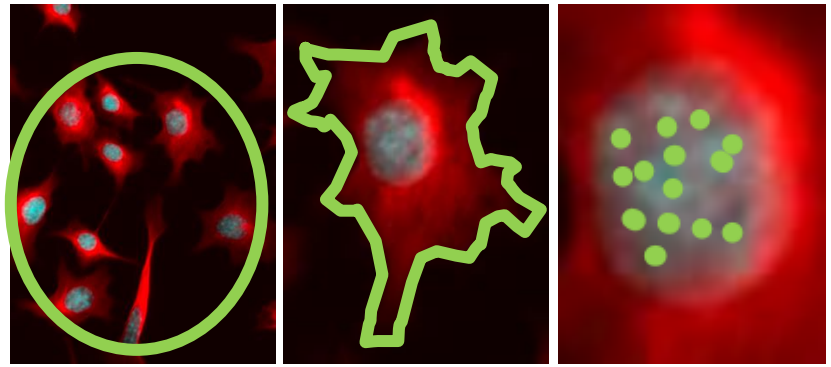
What imaging is good for

Spatial analysis



- morphology
- neighborhood relationships
- multicellular structure
- alignment
- tissue organisation
- motility (w/ time analysis)

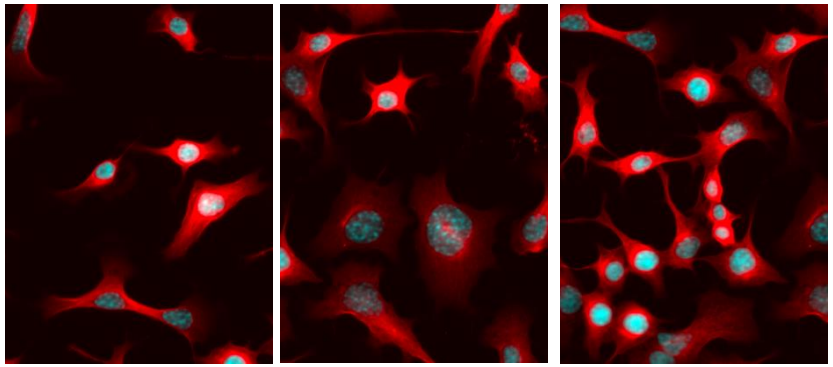
Choice of object scale



cell group cell spots in cell

- relationships between scales
- properties of non-cell objects

Time analysis



0 h 6 h 12 h

- dynamic gene expression
- motility
- properties over time
- single cell dynamics (tracking)

What imaging is not so good for

Concentration

- absolute intensities are often ambiguous
- fluorescence imaging is often indirect

Alternatives:
- qPCR
- RNA sequencing
- ELISA

Throughput

- more resolution
  = less cells per FOV
- more channels
  = longer acquisition time

Alternatives:
- Flow cytometry

Molecular depth

- often max. 5 channels
- sequential staining is difficult

Alternatives:
- RNA sequencing
- Mass spectrometry

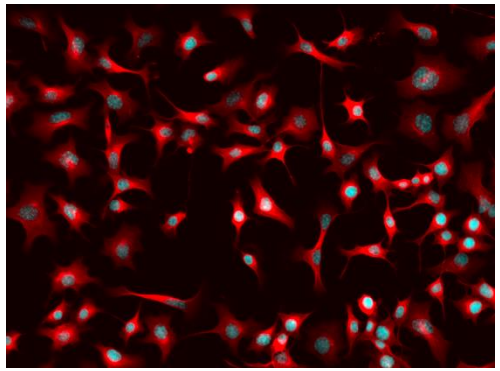
Molecular identity

- fluorescence is an indirect readout
- ABs / probes can bind non-specifically

Alternatives:
- RNA sequencing
- Mass spectrometry
- qPCR

Things to be cautious about

image



segmentation

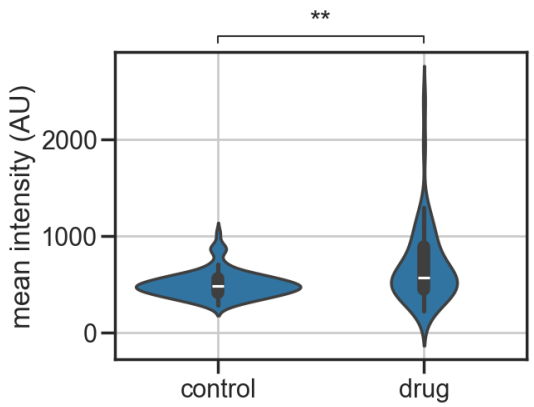
measurements

data

	label	area	mean_dapi	mean_protein
0	1	889.0	0.438285	0.253473
1	2	1245.0	0.357890	0.331311
2	3	609.0	0.448923	0.352874
3	4	1005.0	0.406975	0.341602
4	5	697.0	0.360550	0.341725
5	6	697.0	0.445584	0.264794

analysis

results



DON'T TRUST

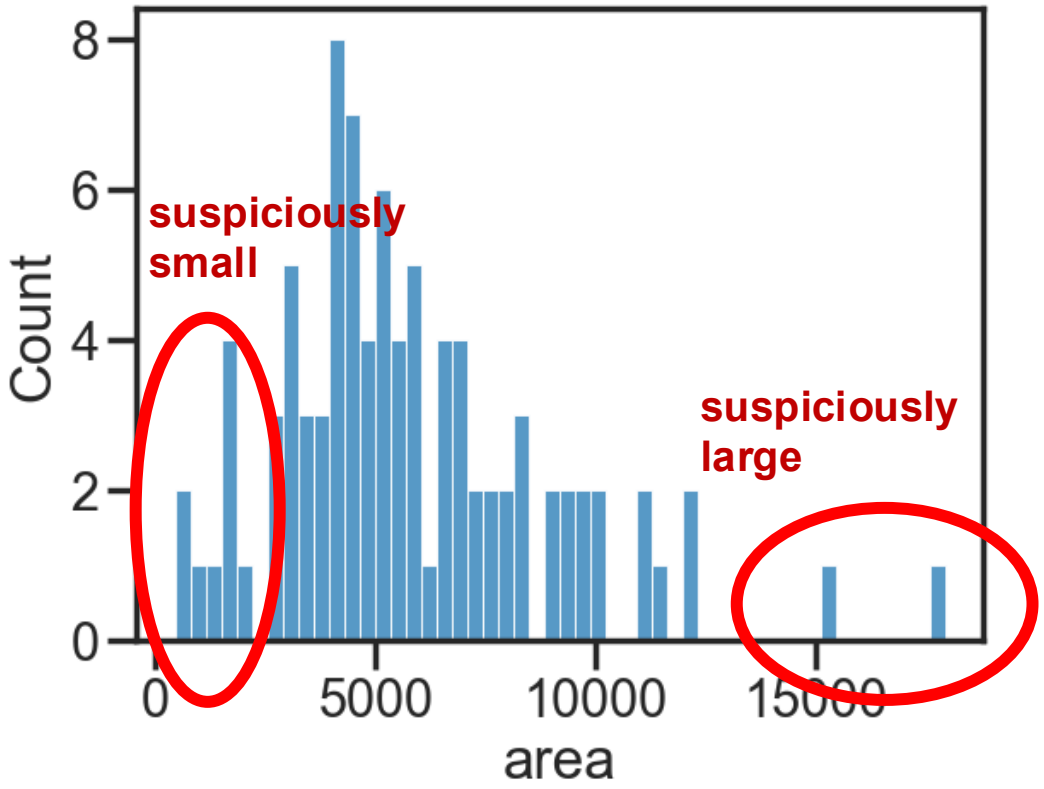
Validate segmentation
Quality control
Validate measurements

DON'T TRUST

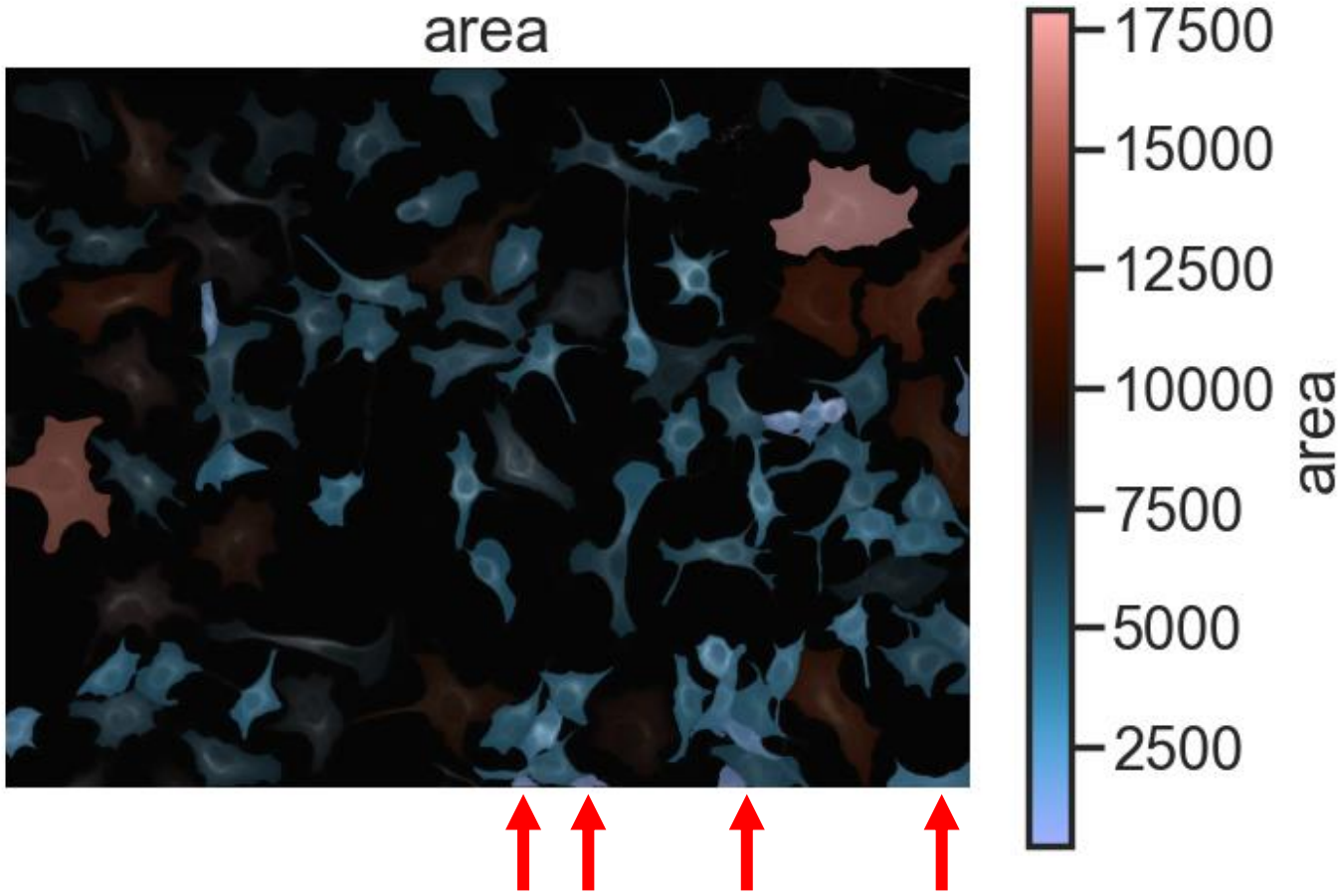
Choose the correct N
Avoid batch effects
Identify confounding measurements

How to spot bad measurements

Check outliers in histograms

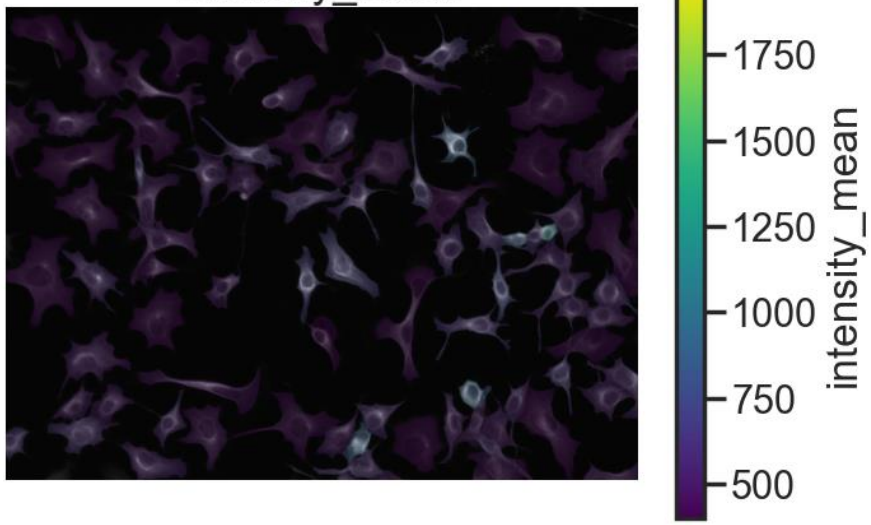


Overlay measurements on image

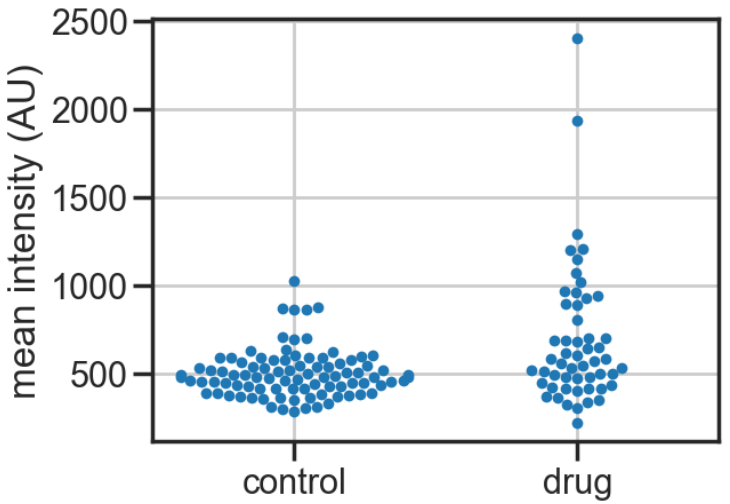


The problem with n and N

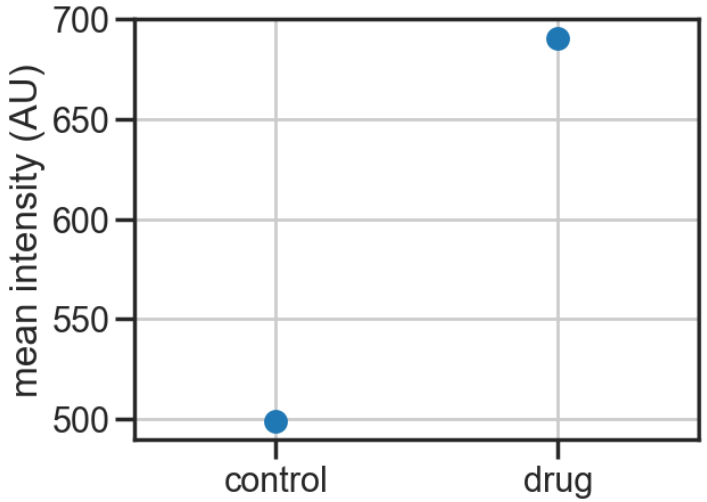
control



Plotting all cells, n



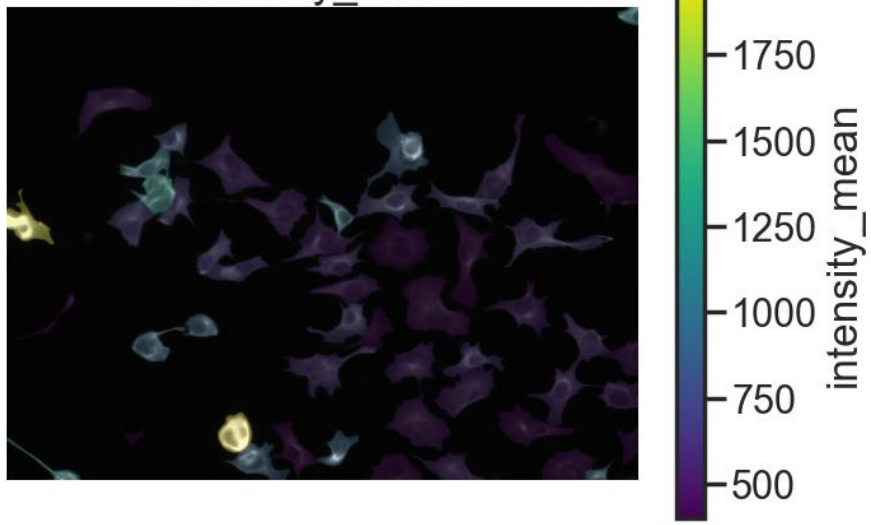
Plotting replicates, N



Individual cells per well are not independent.

This will artificially inflate your statistical power (lower p-value).

drug



Use population averages instead. (of course, we need more than one replicate)

A typical imaging experiment

QUESTION:
Does my drug increase the expression of my gene of interest?

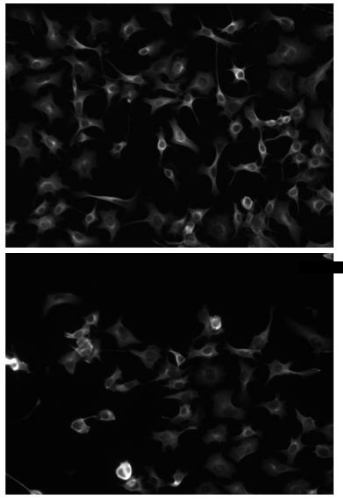
Setup

control

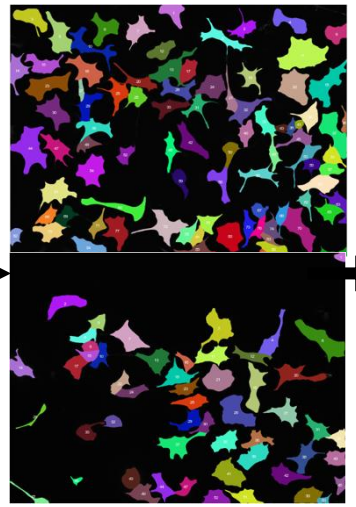
drug

Microscopy

Images



Segmentation

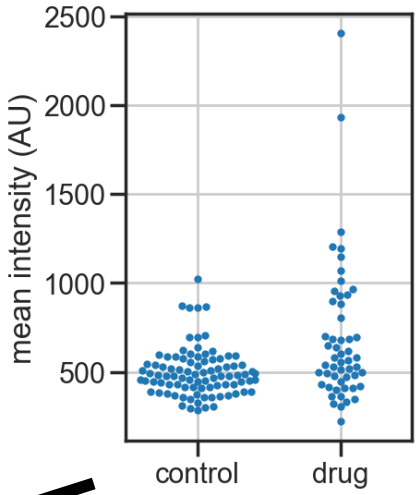


Data

	cell_id	condition	intensity_mean
0	1	control	382.040588
1	2	control	617.344185
2	3	control	283.982710
3	4	control	386.802736
4	5	control	427.812532
...	...	...	...
138	49	drug	926.618473
139	50	drug	410.990932
140	51	drug	528.766274
141	52	drug	363.509784
142	53	drug	957.609105

143 rows × 3 columns

Results



CONCLUSION:
Maybe a little bit.

Questions?