

Streamlining image analysis  
decision-making  
for efficient and unbiased  
biomedical research

# Motivation I

- **Academic:**
  - contribute to practices and platforms that sustain **open, reproducible, collaborative** research and **management** of data

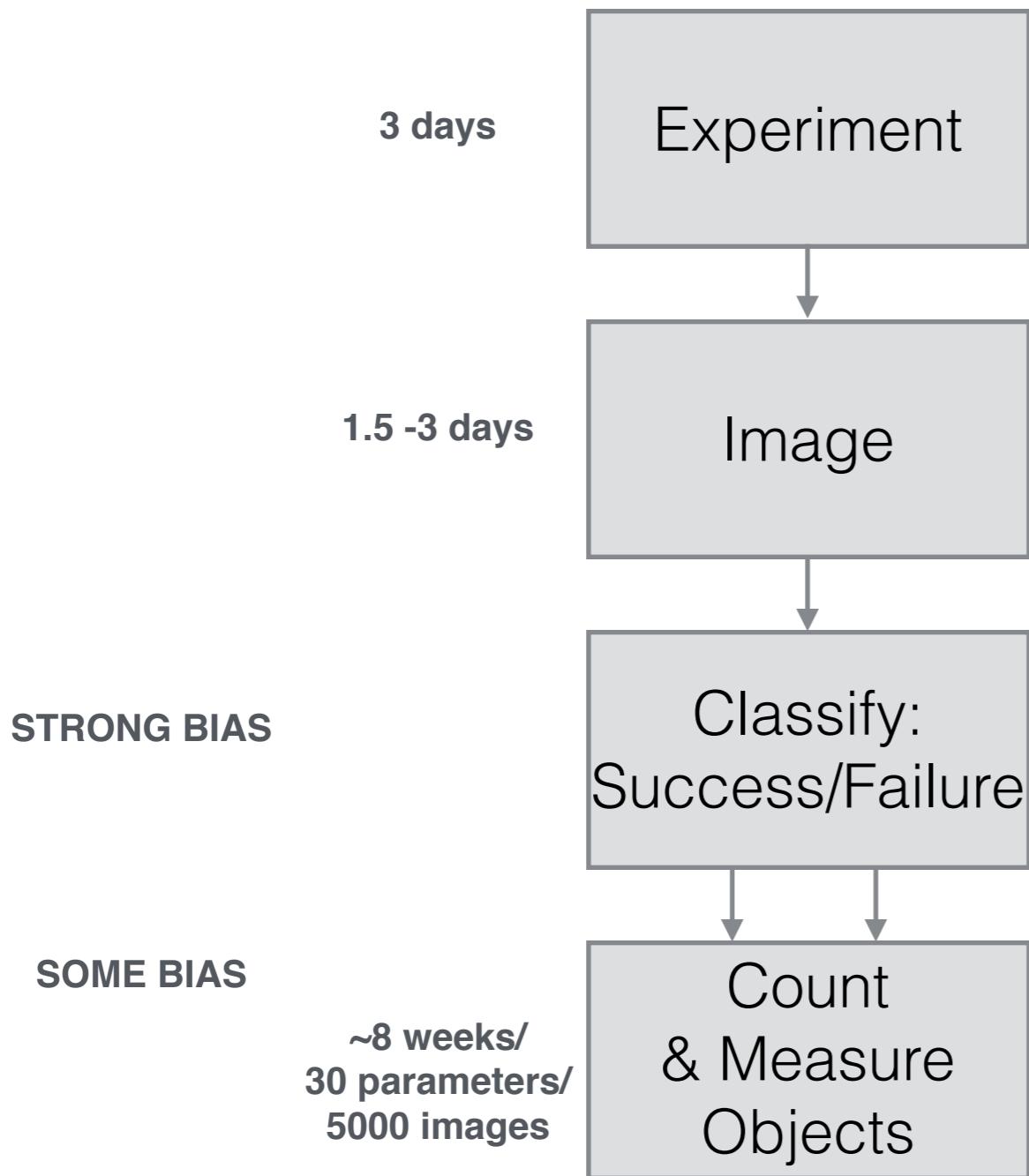
# Motivation II

- **Academic:**
  - contribute to practices and platforms that sustain **open**, **reproducible**, collaborative research and **management** of data
- **Business I:** Decrease long-term variable cost
  - enable researchers to spend more time focusing on their science

# Motivation II

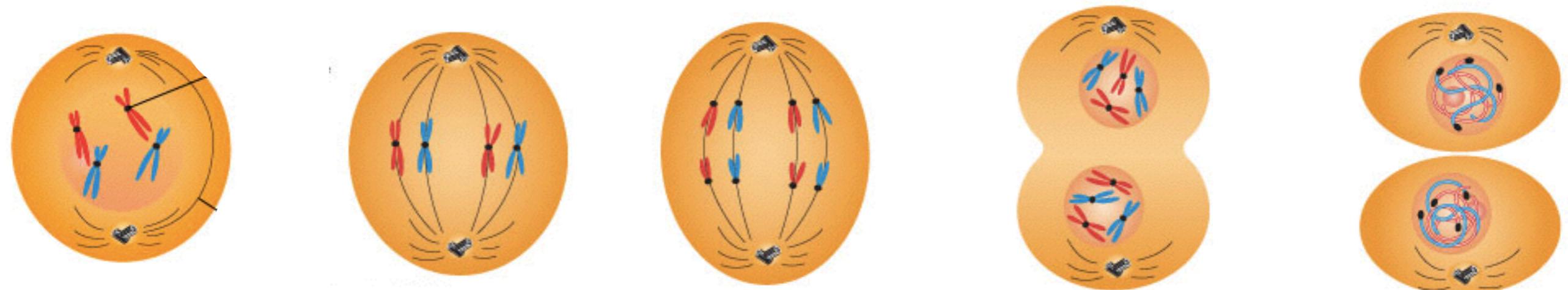
- **Academic:**
  - contribute to practices and platforms that sustain **open**, **reproducible**, collaborative research and **management** of data
- **Business I:** Decrease long-term variable cost
  - enable researchers to spend more time focusing on their science
- **Business I:** Biopharmaceutical ROI problem
  - Save time, eliminate bias, obtain accurate and reproducible numbers

# Problems with the traditional workflow:

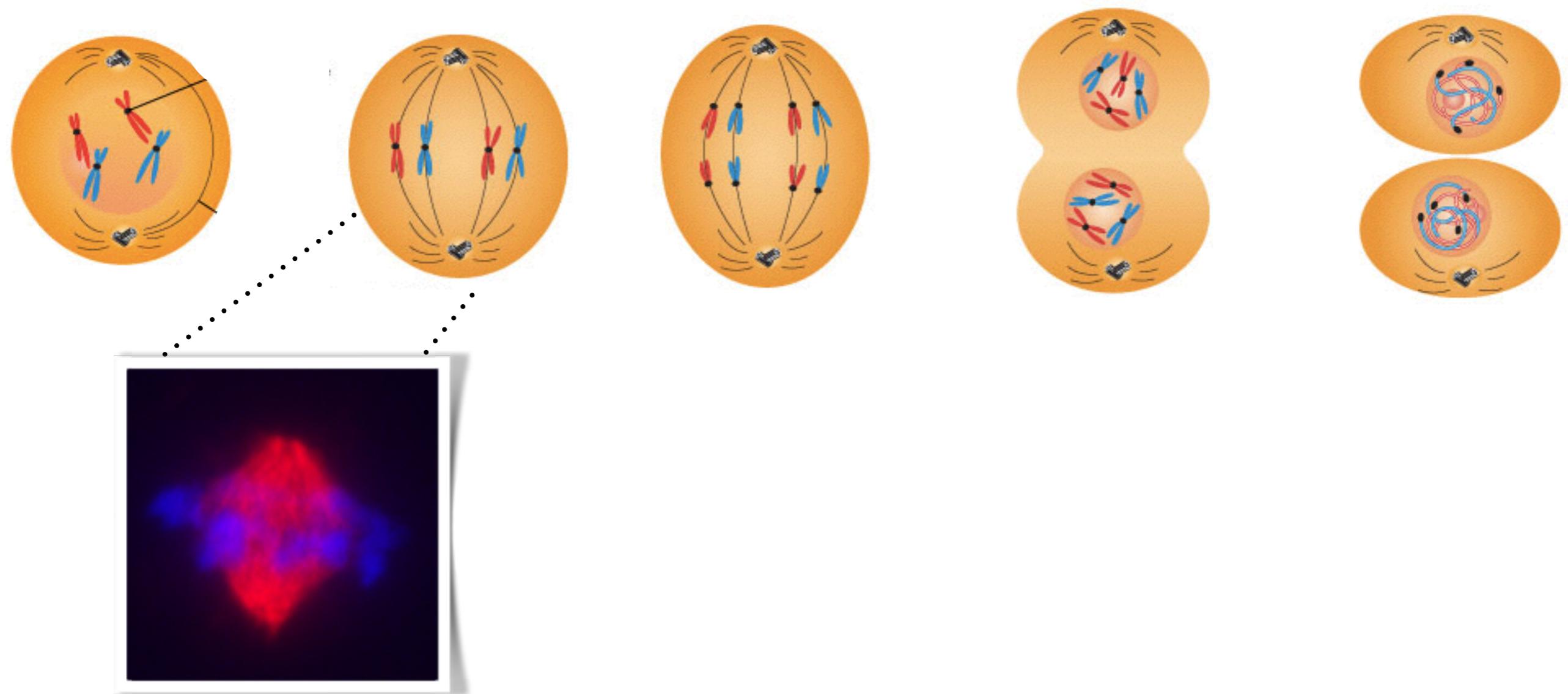


# Case study

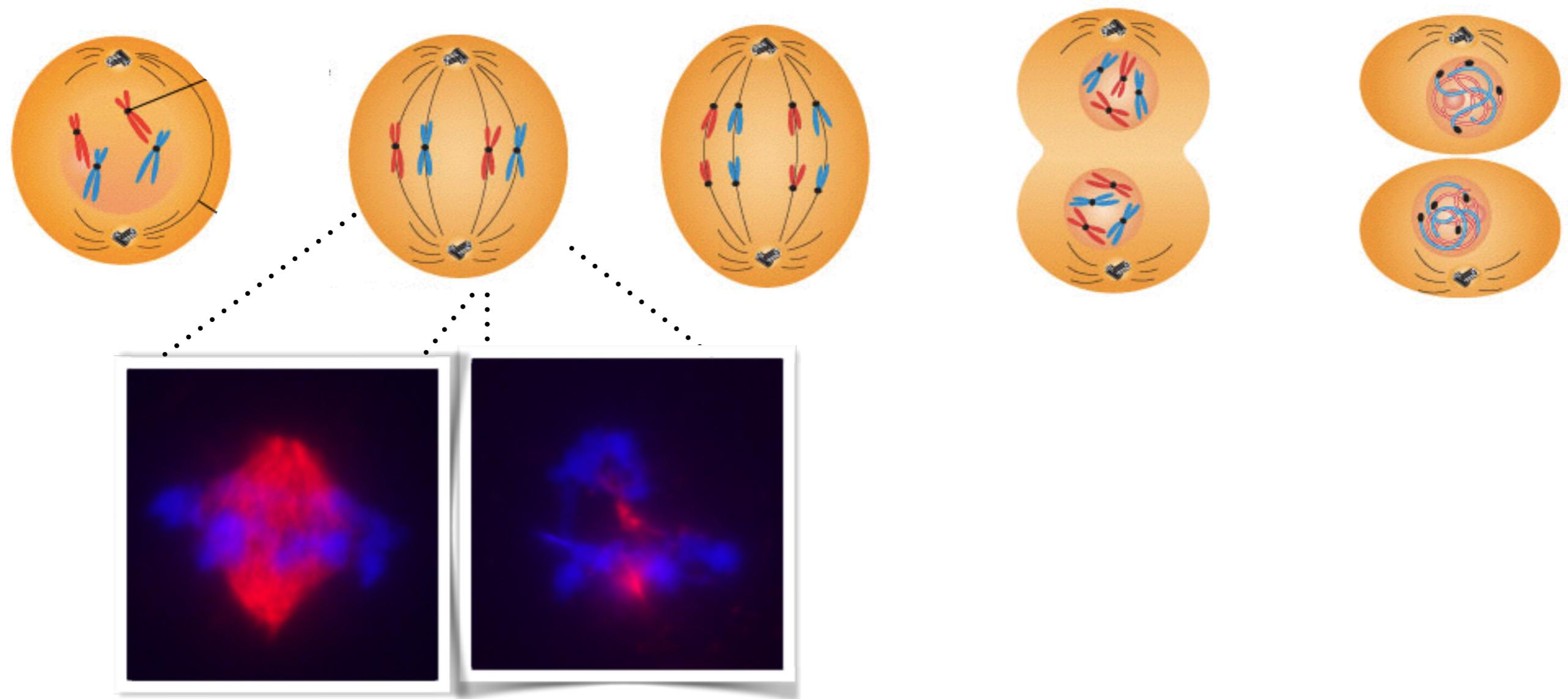
# Comments on the science



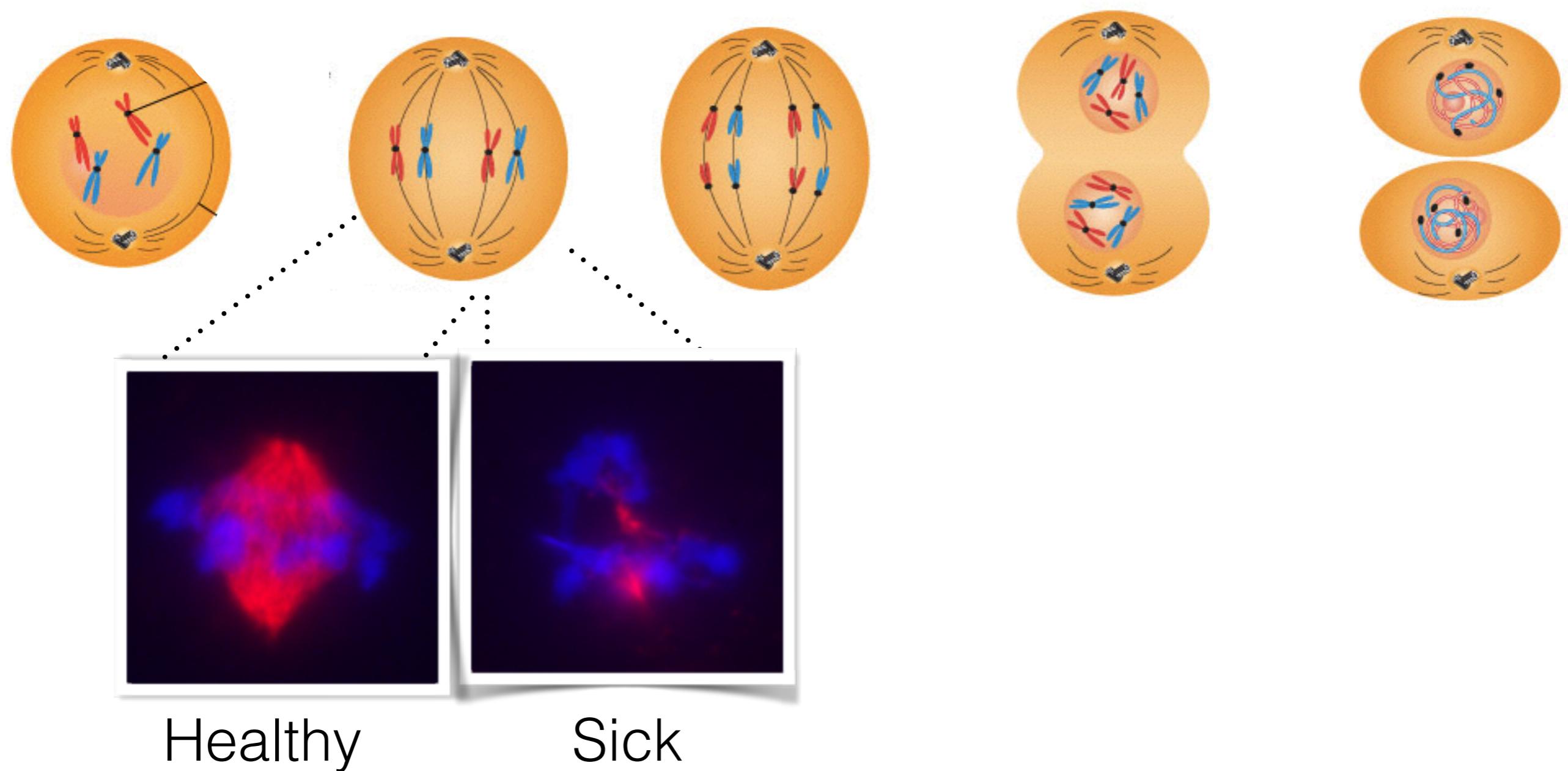
# Comments on the science



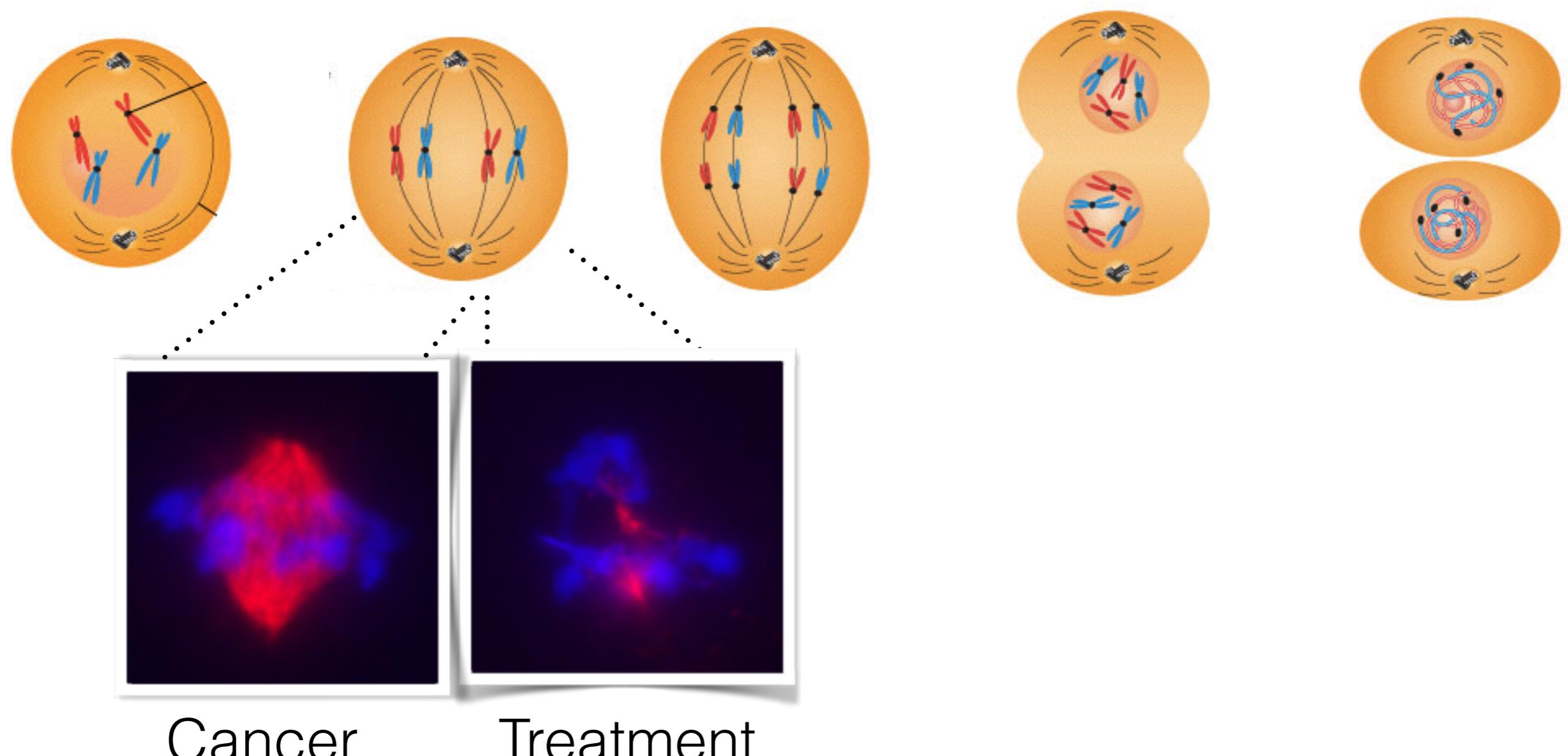
# Comments on the science



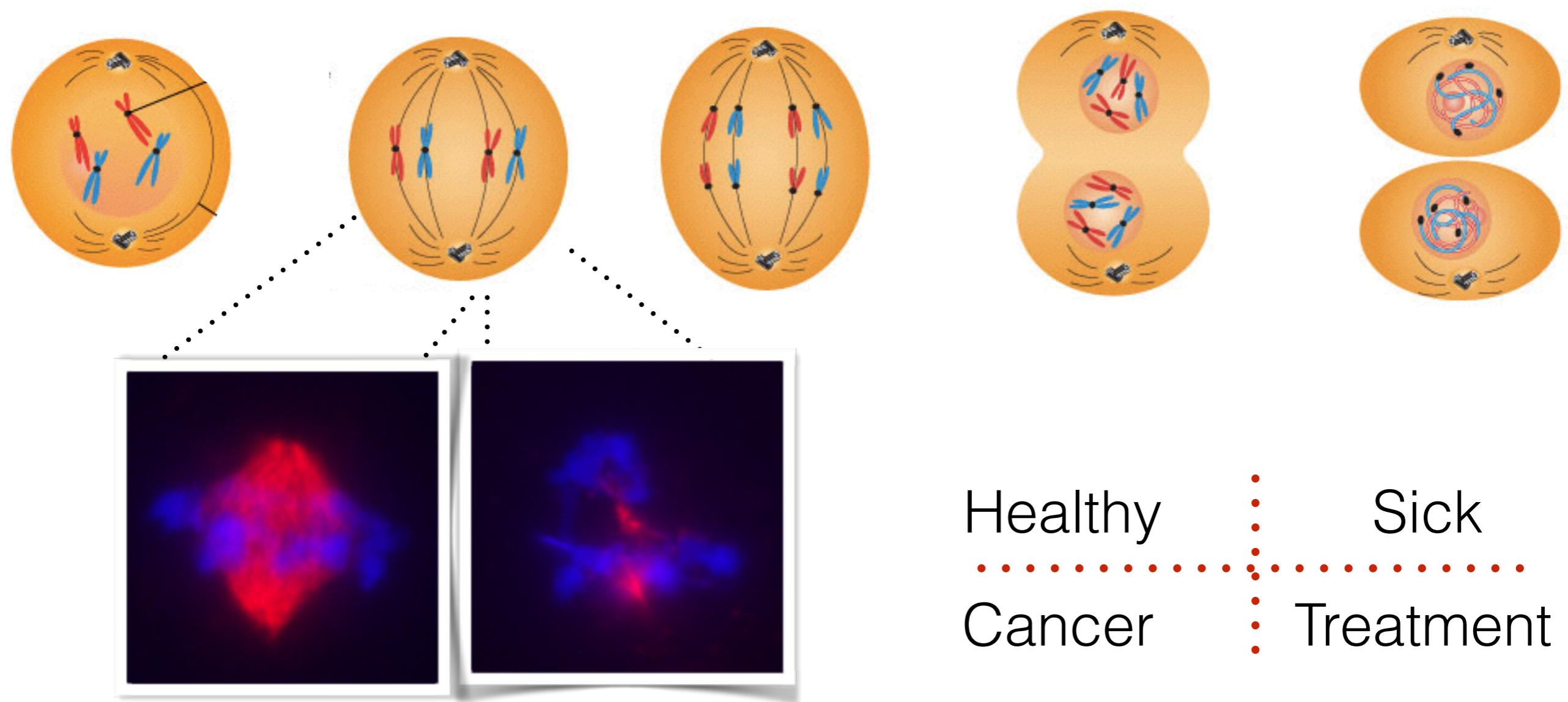
# Comments on the science



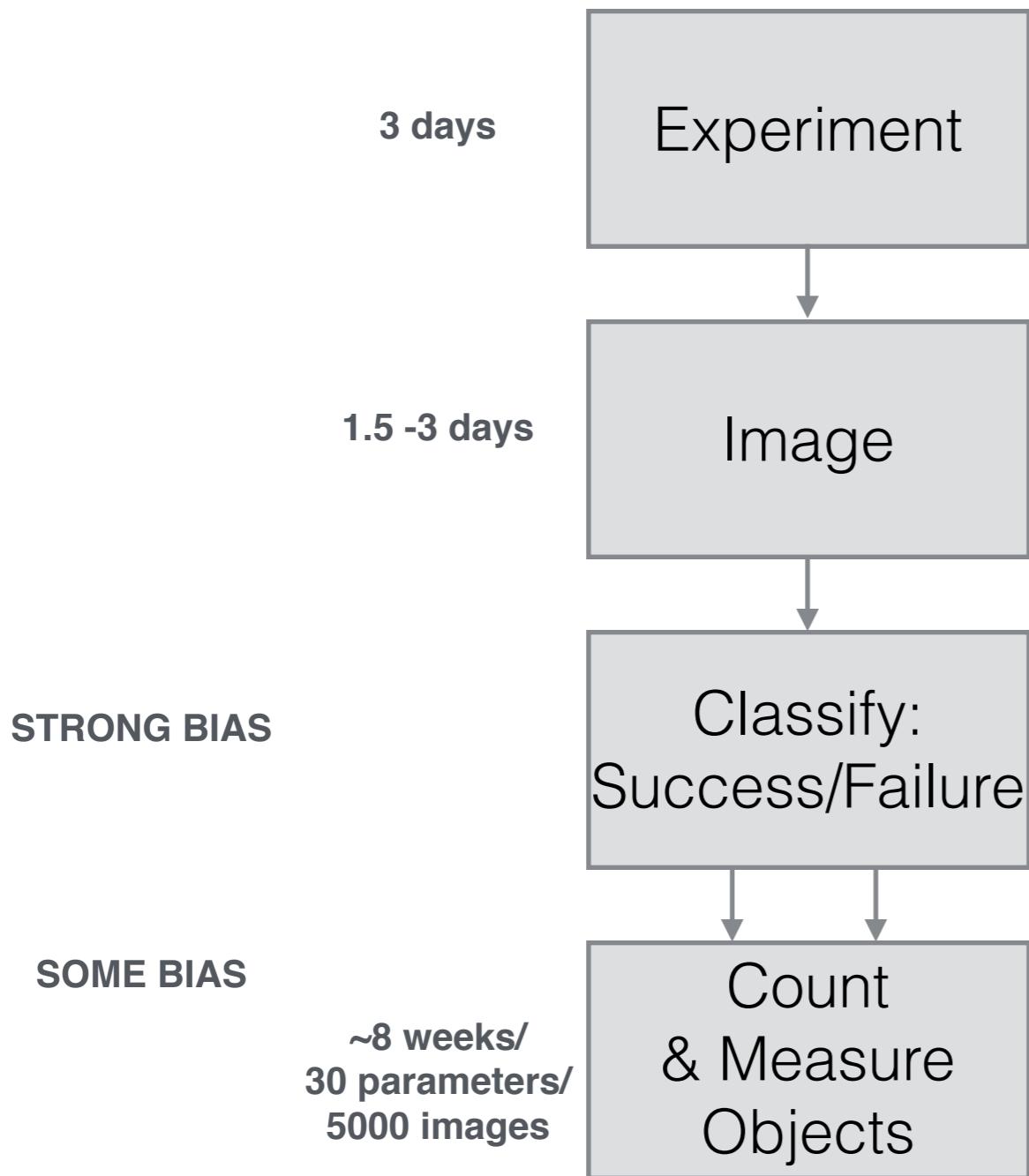
# Comments on the science



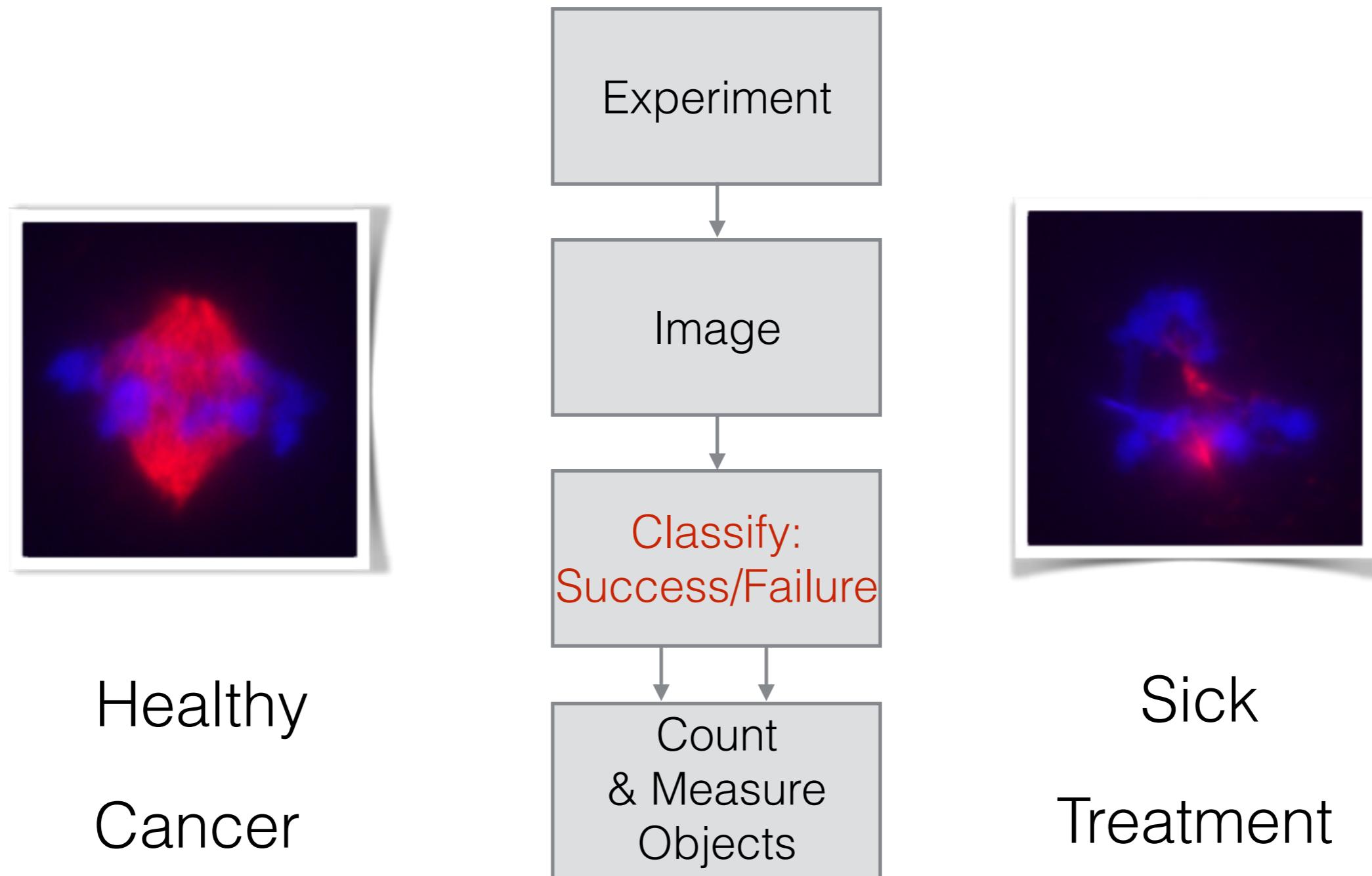
# Comments on the science



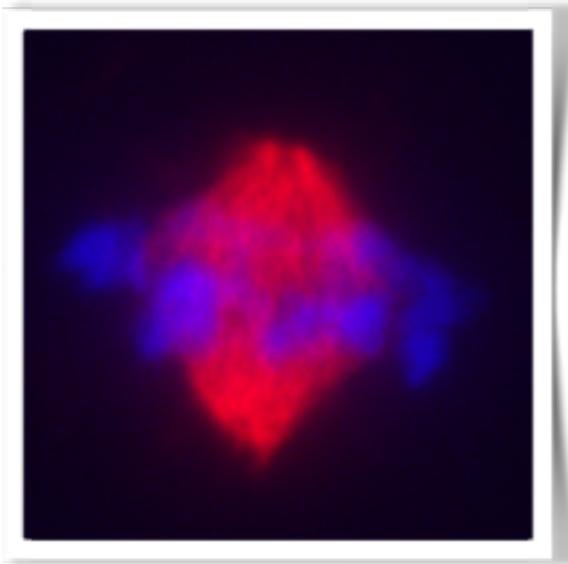
# Problems with the traditional workflow:



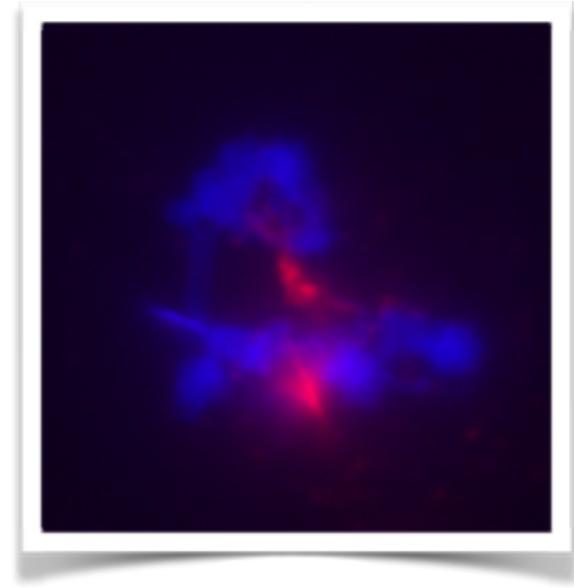
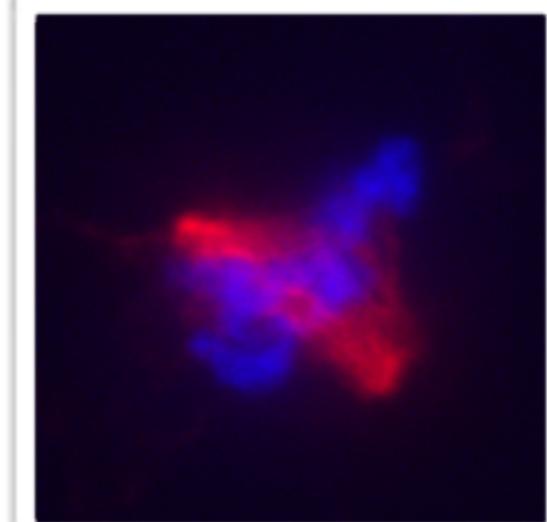
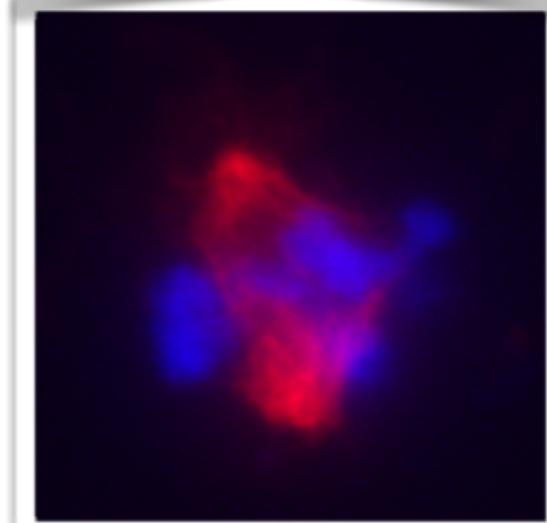
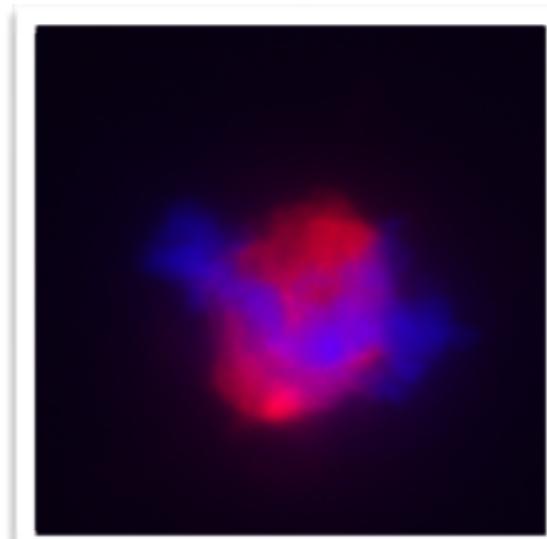
# So why is there a problem?



# So why is there a problem?

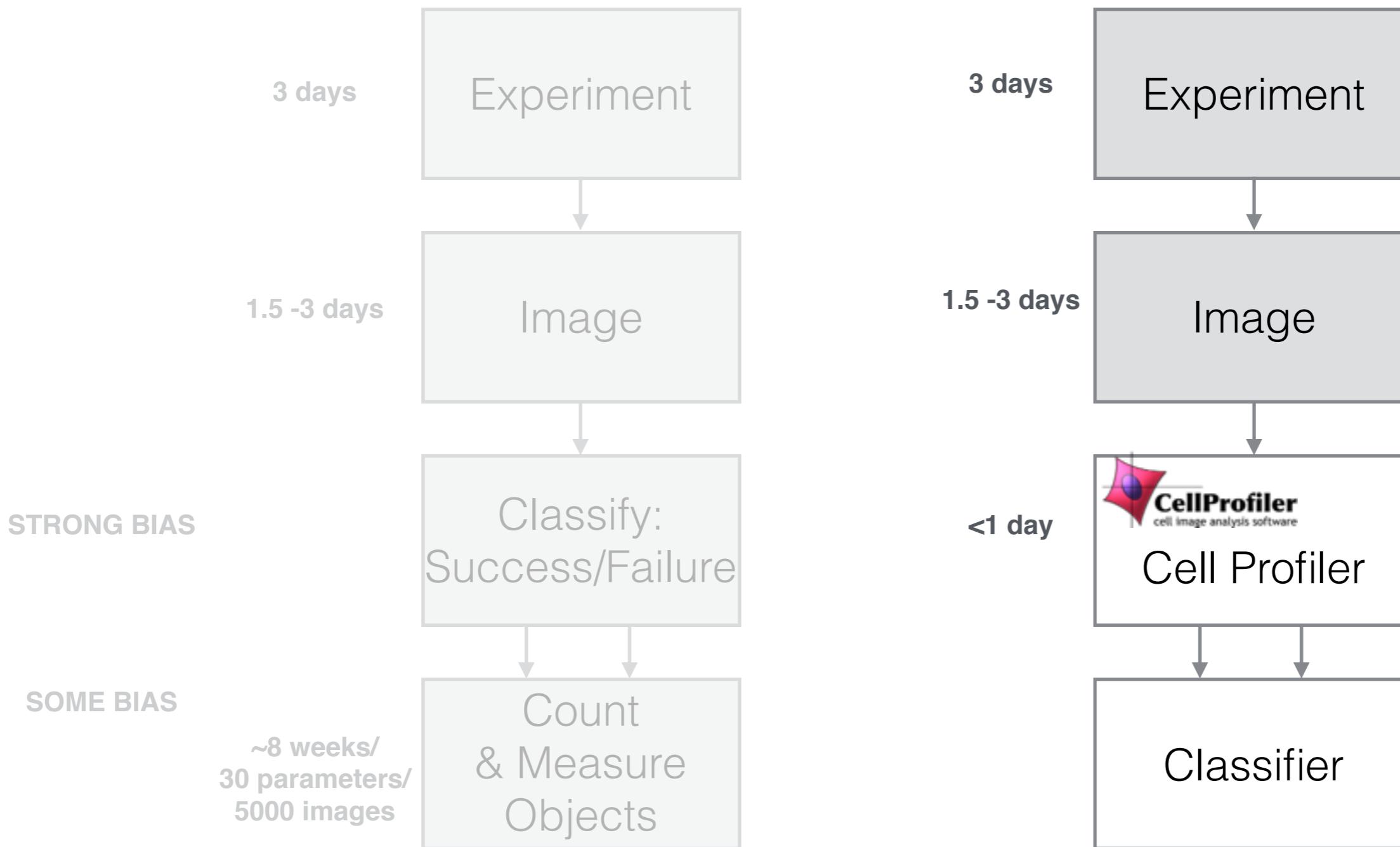


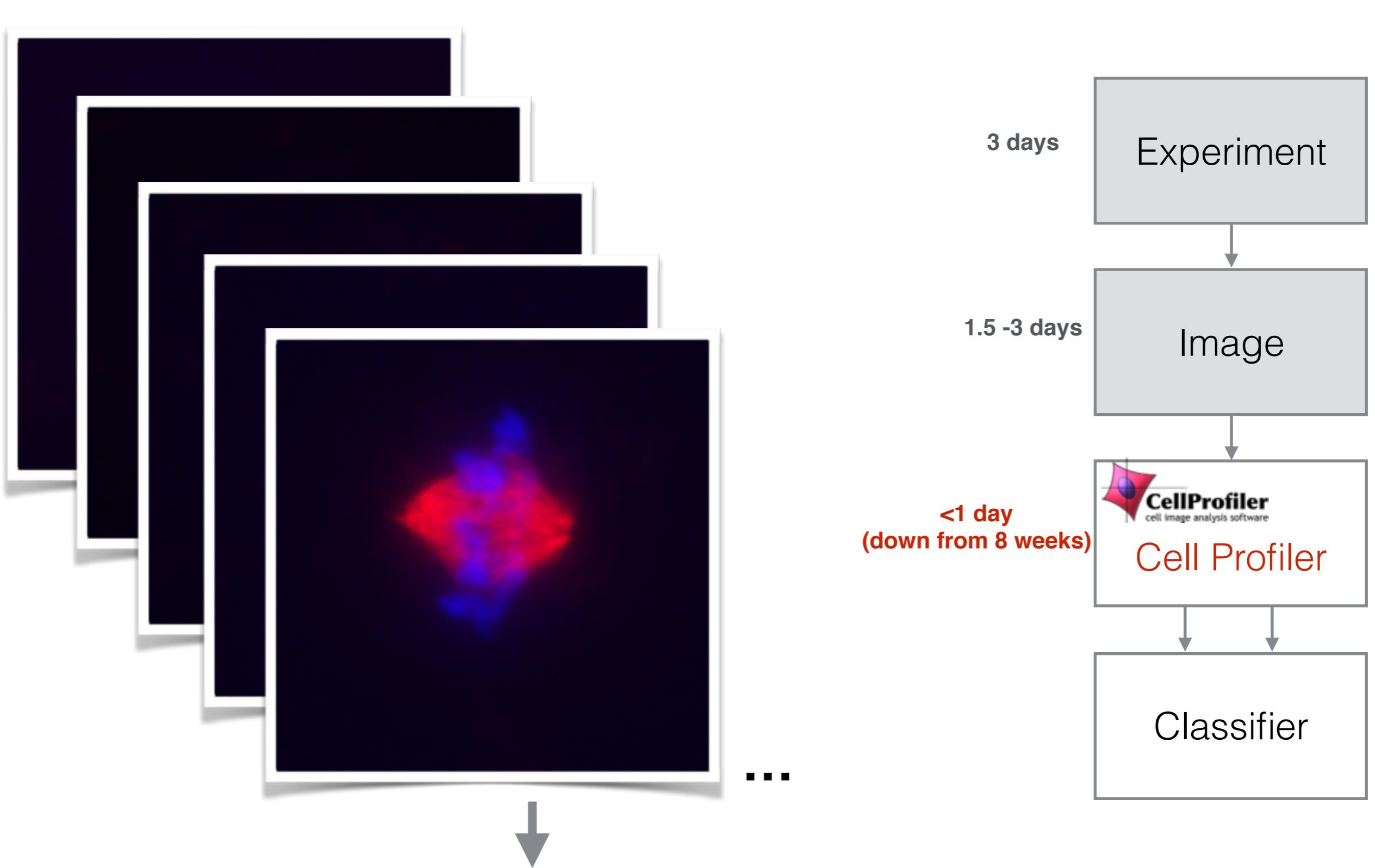
Healthy  
Cancer



Sick  
Treatment

# Problems with the traditional workflow: solution





Y	TREATMENT	CONC_TREATMENT	SOLVENT	CONC_SOLVENT	...	Texture_Difference	Variance_CorrMT_3_0	Texture_Entropy_CorrMT_3_0	Texture_Gabor
0	1	0	0	0	...	0.297365		2.684328	0.409684
0	1	0	0	0	...	0.291719		2.767921	0.740311
0	1	0	0	0	...	0.279736		2.734776	0.369790
0	1	0	0	0	...	0.322350		2.996500	0.519450
0	1	0	0	0	...	0.338992		2.902682	0.888699

# Data Set

$x^{(j)}$  500 features (106 here)

...

Y	TREATMENT	CONC_TREATMENT	SOLVENT	CONC_SOLVENT	...	Texture_DifferenceVariance_CorrMT_3_0	Texture_Entropy_CorrMT_3_0	Texture_Gaboi
0	1	0	0	0	...	0.297365	2.684328	0.409684
0	1	0	0	0	...	0.291719	2.767921	0.740311
0	1	0	0	0	...	0.279736	2.734776	0.369790
0	1	0	0	0	...	0.322350	2.996500	0.519450
0	1	0	0	0	...	0.338992	2.902682	0.888699

:  $x^{(i)}$

: 13000 examples (901 here)

**Features:** shape, texture, intensity, etc.  
**Examples:** quantified images

Y	TREATMENT	CONC_TREATMENT	SOLVENT	CONC_SOLVENT	...	Texture_DifferenceVariance_CorrMT_3_0	Texture_Entropy_CorrMT_3_0	Texture_Gabor
0	1	0	0	0	...	0.297365	2.684328	0.409684
0	1	0	0	0	...	0.291719	2.767921	0.740311
0	1	0	0	0	...	0.279736	2.734776	0.369790
0	1	0	0	0	...	0.322350	2.996500	0.519450
0	1	0	0	0	...	0.338992	2.902682	0.888699

3 days

Experiment

1.5 -3 days

Image

<1 day  
(down from 8 weeks)

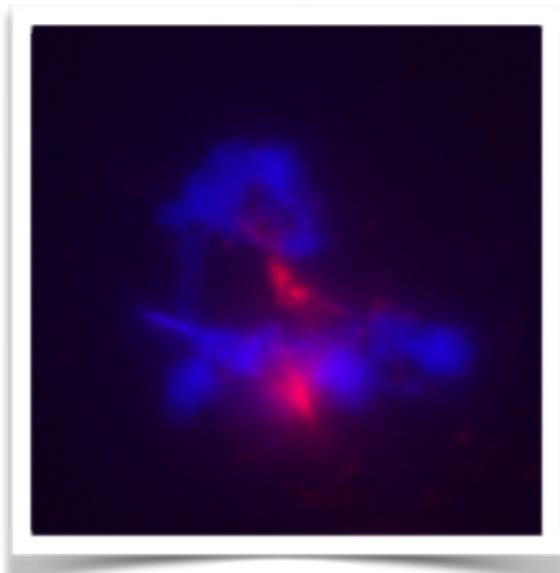
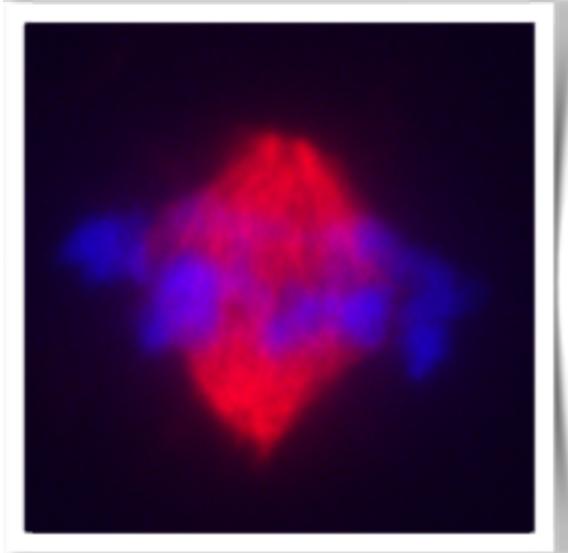


Cell Profiler

Classifier

Healthy

Sick



Y	TREATMENT	CONC_TREATMENT	SOLVENT	CONC_SOLVENT	...	Texture_DifferenceVariance_CorrMT_3_0	Texture_Entropy_CorrMT_3_0	Texture_Gaboi
0	1	0	0	0	...	0.297365	2.684328	0.409684
0	1	0	0	0	...	0.291719	2.767921	0.740311
0	1	0	0	0	...	0.279736	2.734776	0.369790
0	1	0	0	0	...	0.322350	2.996500	0.519450
0	1	0	0	0	...	0.338992	2.902682	0.888699

3 days

Experiment

1.5 -3 days

Image

<1 day  
(down from 8 weeks)

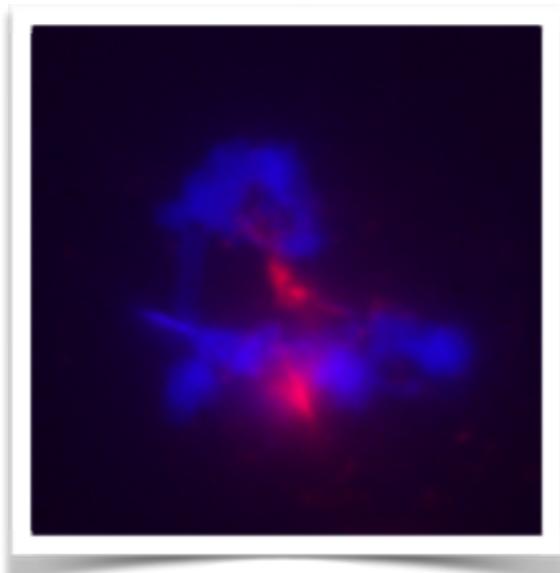
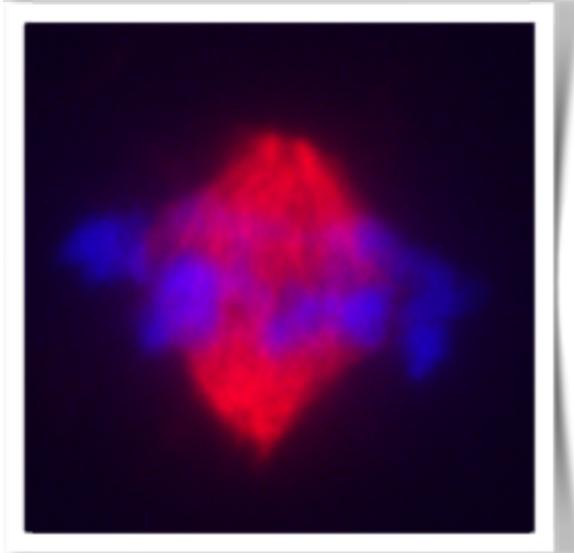


Cell Profiler

Classifier

Healthy

Sick

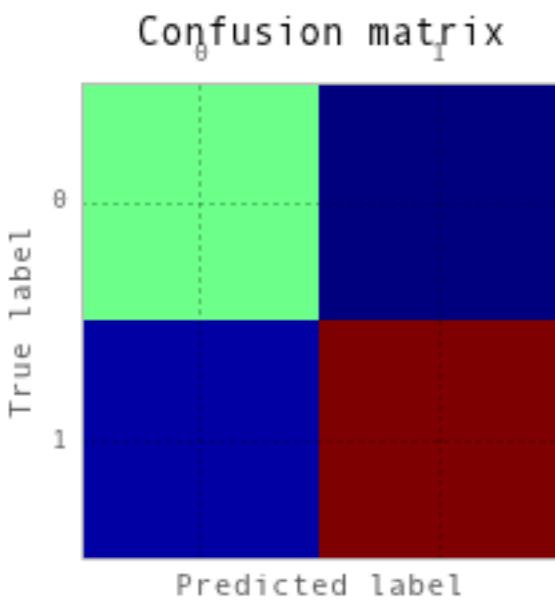


Logistic regression  
Decision Trees

Random Forests  
SVMs  
Neural Network

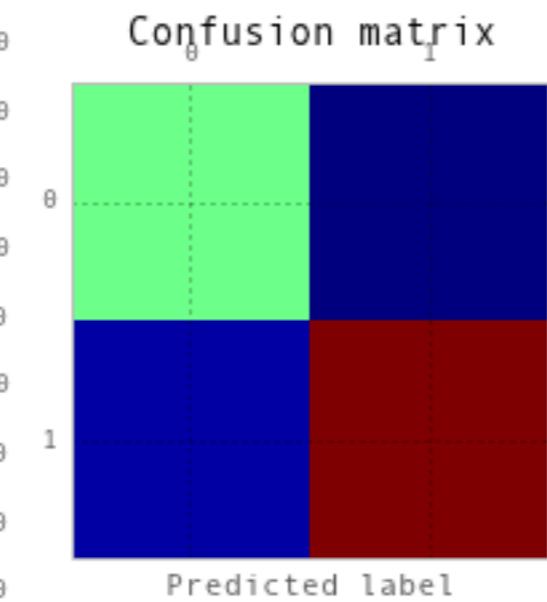
# Summary

Logistic regression



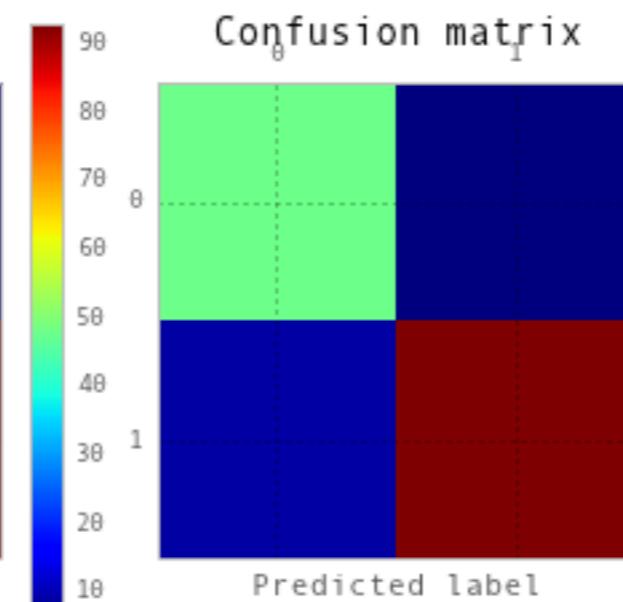
F1 score:  
0.90

SVM (RBF)



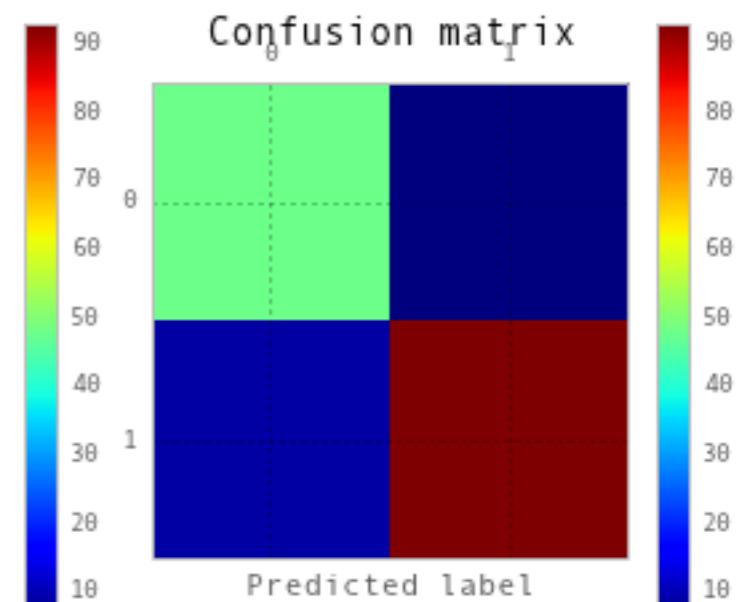
F1 score:  
0.93

Decision Tree



F1 score:  
0.81

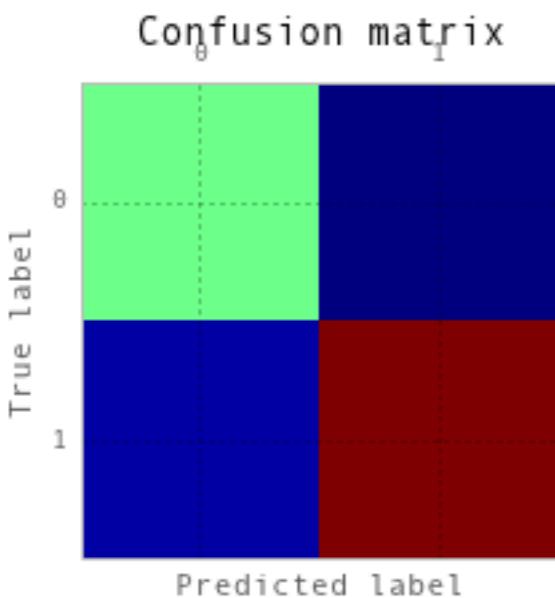
Random forest



F1 score:  
0.87

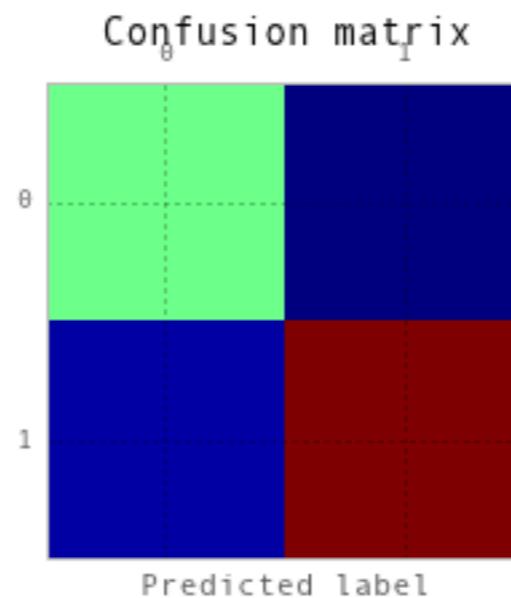
# Summary

## Logistic regression



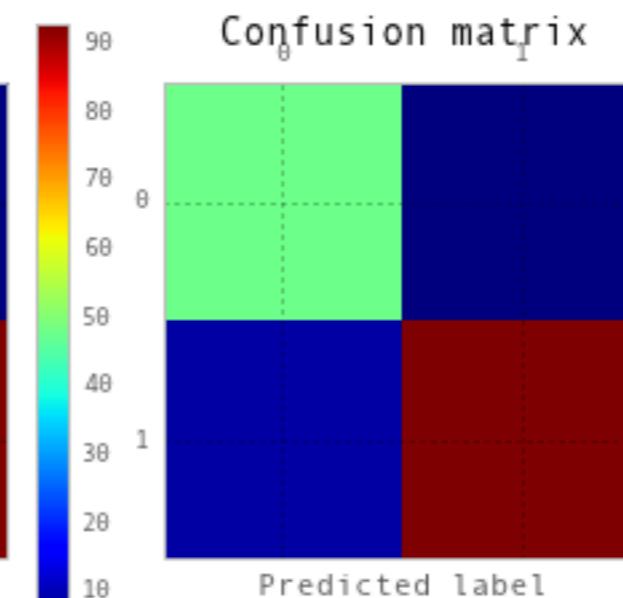
F1 score:  
0.90

## SVM (RBF)



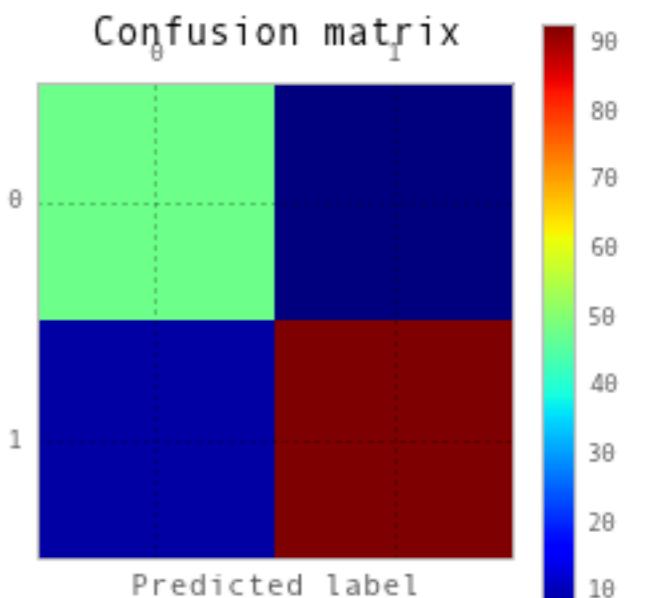
F1 score:  
0.93

## Decision Tree



F1 score:  
0.81

## Random forest



F1 score:  
0.87

$x^{(j)}$  500 features (106 here)

...

Y	TREATMENT	CONC_TREATMENT	SOLVENT	CONC_SOLVENT	...	Texture_DifferenceVariance_CorrMT_3_0	Texture_Entropy_CorrMT_3_0	Texture_Gaboi
0	1	0	0	0	...	0.297365	2.684328	0.409684
0	1	0	0	0	...	0.291719	2.767921	0.740311
0	1	0	0	0	...	0.279736	2.734776	0.369790
0	1	0	0	0	...	0.322350	2.996500	0.519450
0	1	0	0	0	...	0.338992	2.902682	0.888699

:  $x^{(i)}$

13000 examples (901 here)

**Features:** shape, texture, intensity, etc.  
**Examples:** quantified images

$x^{(j)}$  500 features (106 here)

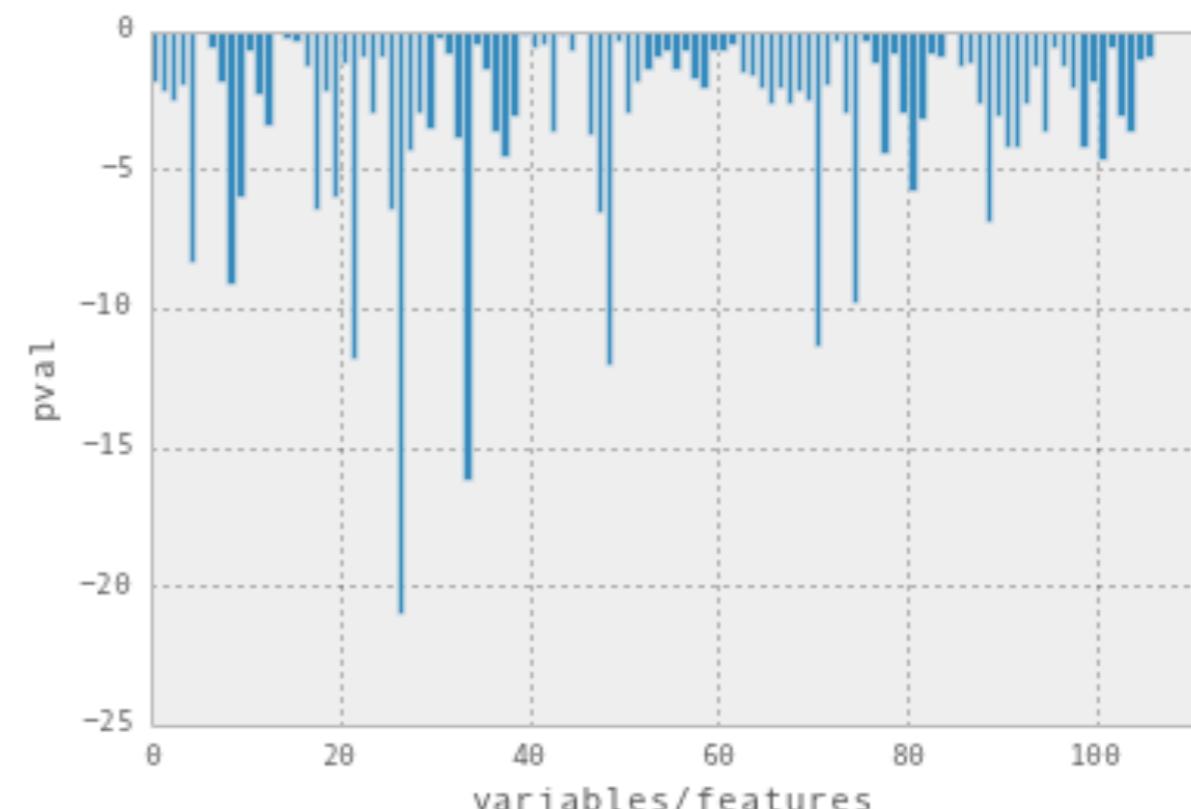
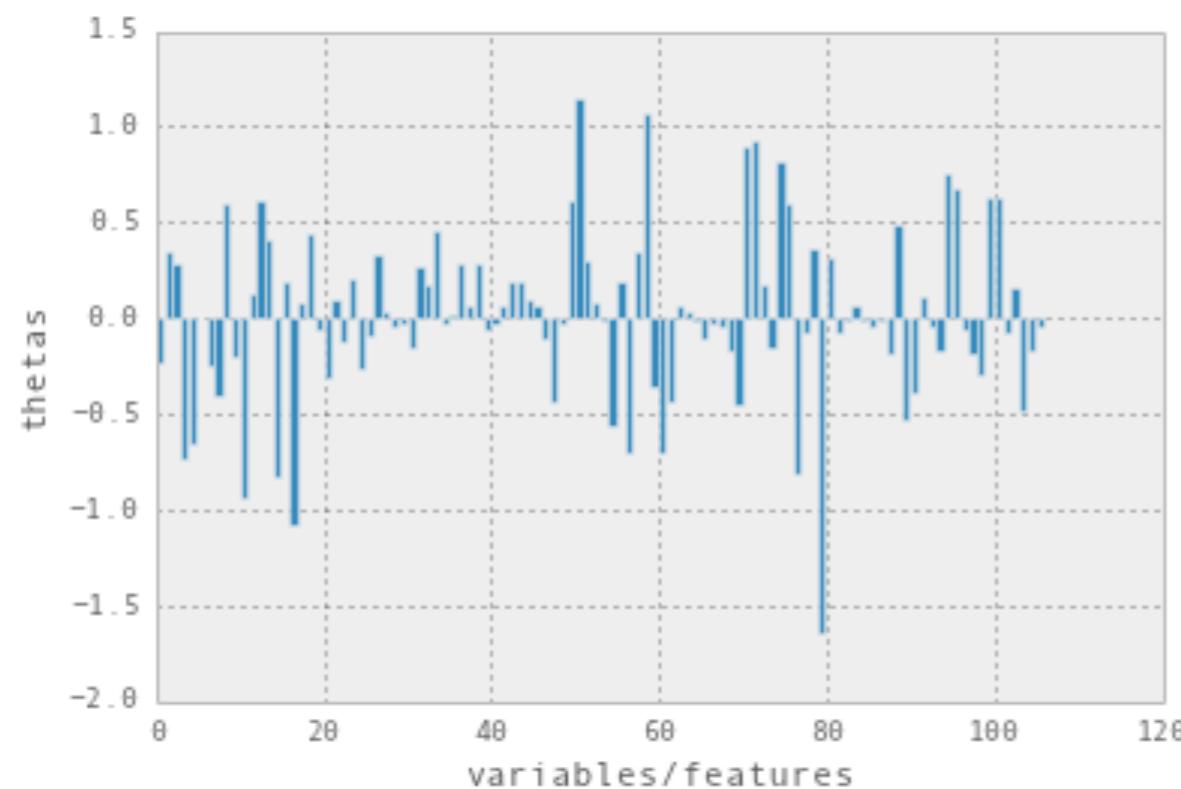
...

Y	TREATMENT	CONC_TREATMENT	SOLVENT	CONC_SOLVENT	...	Texture_DifferenceVariance_CorrMT_3_0	Texture_Entropy_CorrMT_3_0	Texture_Gaboi
0	1	0	0	0	...	0.297365	2.684328	0.409684
0	1	0	0	0	...	0.291719	2.767921	0.740311
0	1	0	0	0	...	0.279736	2.734776	0.369790
0	1	0	0	0	...	0.322350	2.996500	0.519450
0	1	0	0	0	...	0.338992	2.902682	0.888699

$x^{(i)}$

13000 examples (901 here)

**Features:** shape, texture, intensity, etc.  
**Examples:** quantified images



# Numbers tell the story...

$x^{(j)}$  500 features (106 here)

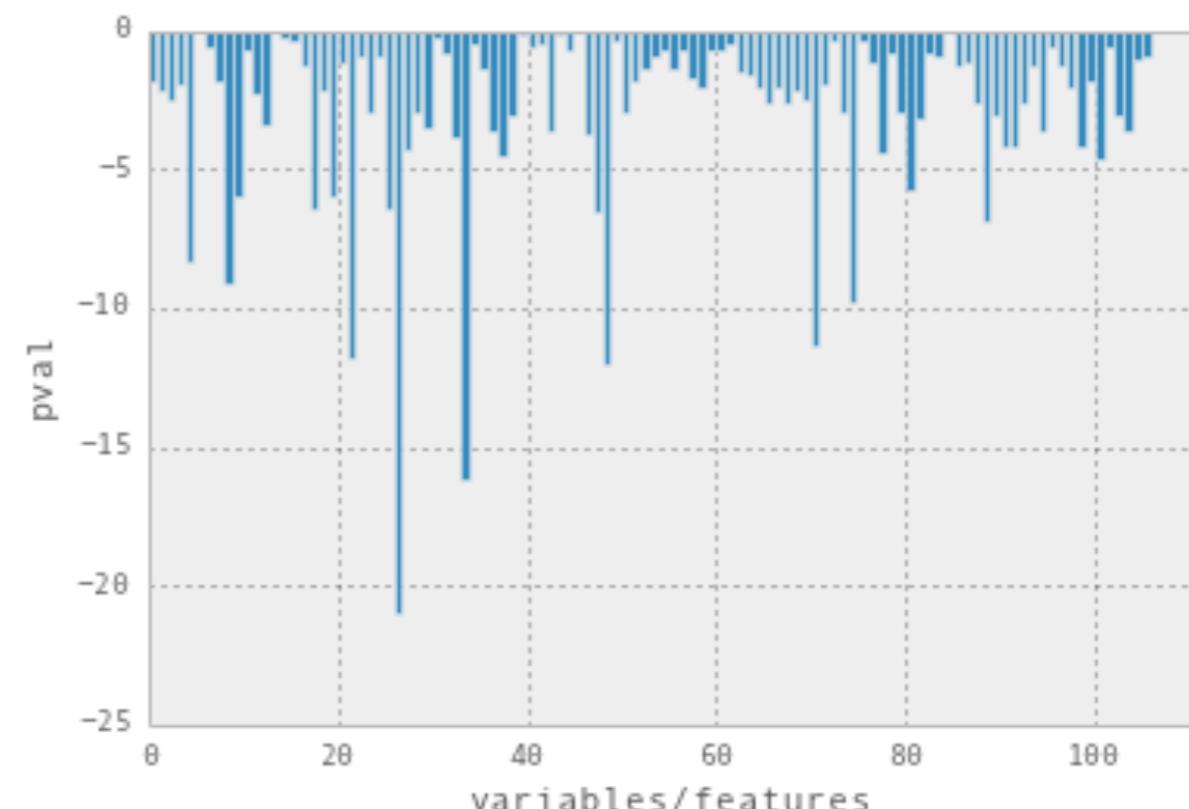
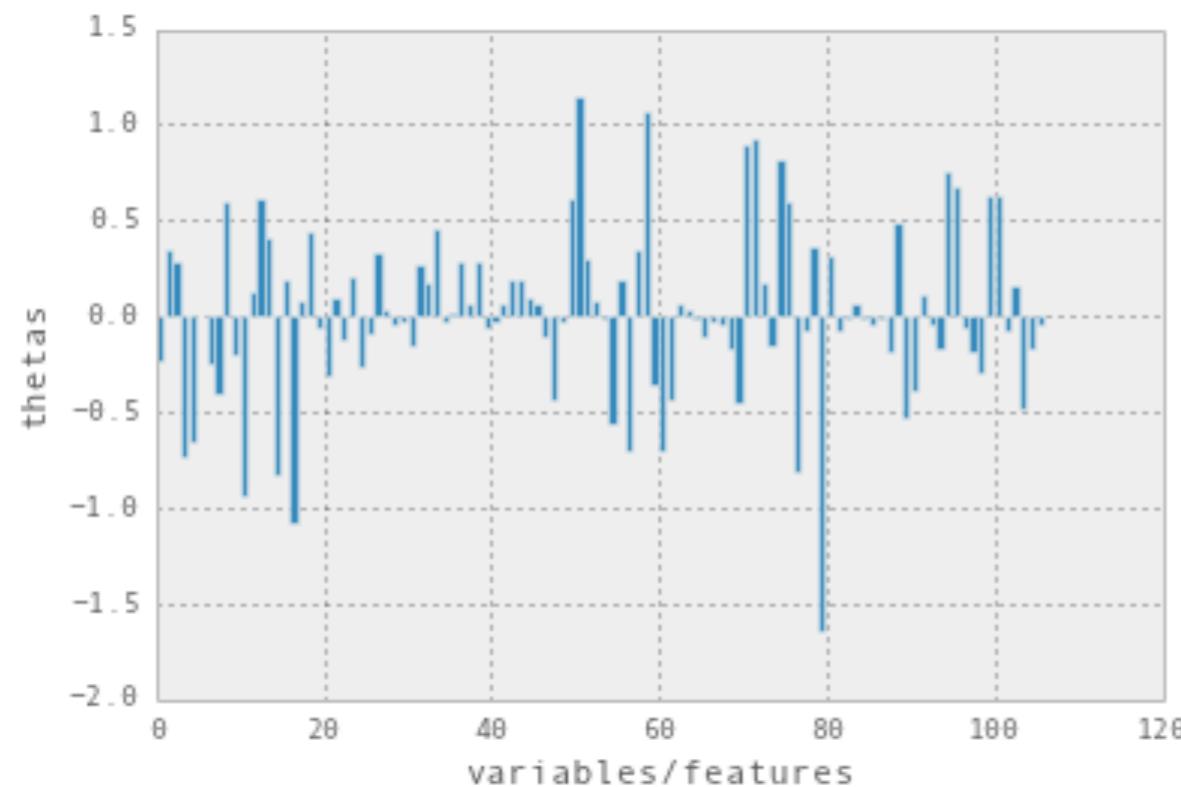
...

Y	TREATMENT	CONC_TREATMENT	SOLVENT	CONC_SOLVENT	...	Texture_DifferenceVariance_CorrMT_3_0	Texture_Entropy_CorrMT_3_0	Texture_Gaboi
0	1	0	0	0	...	0.297365	2.684328	0.409684
0	1	0	0	0	...	0.291719	2.767921	0.740311
0	1	0	0	0	...	0.279736	2.734776	0.369790
0	1	0	0	0	...	0.322350	2.996500	0.519450
0	1	0	0	0	...	0.338992	2.902682	0.888699

$x^{(i)}$

13000 examples (901 here)

**Features:** shape, texture, intensity, etc.  
**Examples:** quantified images



# Summary: Advantages

- Automated multivariate quantification
- Automated subtype classification (max F-1 ~0.93)
- Automated feature (variable selection)
- Easier, comprehensive narrative

# Next steps:

- Increase training set size
- Technical tweaks (learning curves, balance data set, use SVD for imputing missing values)
- Run classification on raw, rather than quantified images

