Molecular interactions investigated by quantitative colocalization measurements and number and brightness (N&B) analysis

Peter Nagy
Dept. of Biophysics and Cell Biology, University of Debrecen

Why is colocalization investigated?

- The interaction of fluorescently labeled, biologically relevant molecules is to be shown.
- The interaction can be studied at different levels (nanometer-micrometer scale), using different methods (e.g. FRET, correlation microscopy, co-precipitation, fluorescence complementation, yeast two-hybrid, etc.).
- Colocalization: being present at the same place.
- How can colocalization be present?

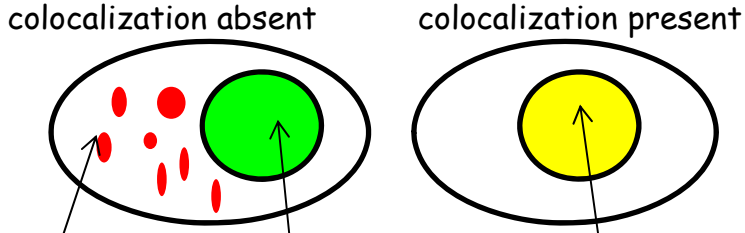
1. **chance** colocalization (e.g. due to overexpression when the overexpressed proteins saturate the sorting systems of a cell and they appear at locations where they normally don't)
2. **apparent** colocalization due to inappropriate methods (see next slide)
3. **real** colocalization where the assumption of direct or indirect interactions between the molecules is correct.

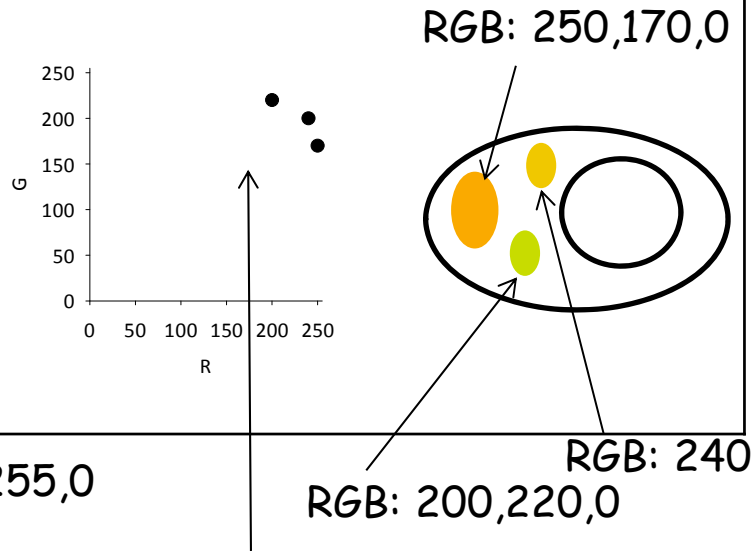
Colocalization measurements are used for the investigation of molecular interactions, BUT
only "real" colocalization implies molecular interactions!

What is colocalization?

Colocalization: in a sample labeled by two different fluorophores

Ordinary definition	**Scientific definition**
if the signals of the two fluorescent labels are at the same place	if the **correlation** between the distributions of the two fluorescent labels is larger than expected for random distributions
E.g. in a sample labeled by green and red fluorophores yellow implies colocalization, but does it?	





RGB: 0,255,0 RGB: 255,255,0

RGB (red, green, blue) code: 255,0,0

Although based on the presence of yellowish color colocalization is assumed, analysis of correlation does not support this assumption.

Quantitative analysis is required.

Methods used for the quantitative analysis of colocalization

ICCB (intensity correlation coefficient-based)	Object-based analysis
• Pearson correlation coefficient • Manders coefficient • Costes' method • van Steensel's method • Li's method	Difficult to classify, not that wide-spread, usually involves relatively complex image analysis

A guided tour into subcellular colocalization analysis in light microscopy
S. Bolte, F.P. Cordelières
J. Microscopy, 224: 213-232 (2006)

ICCB 1: Pearson correlation coefficient

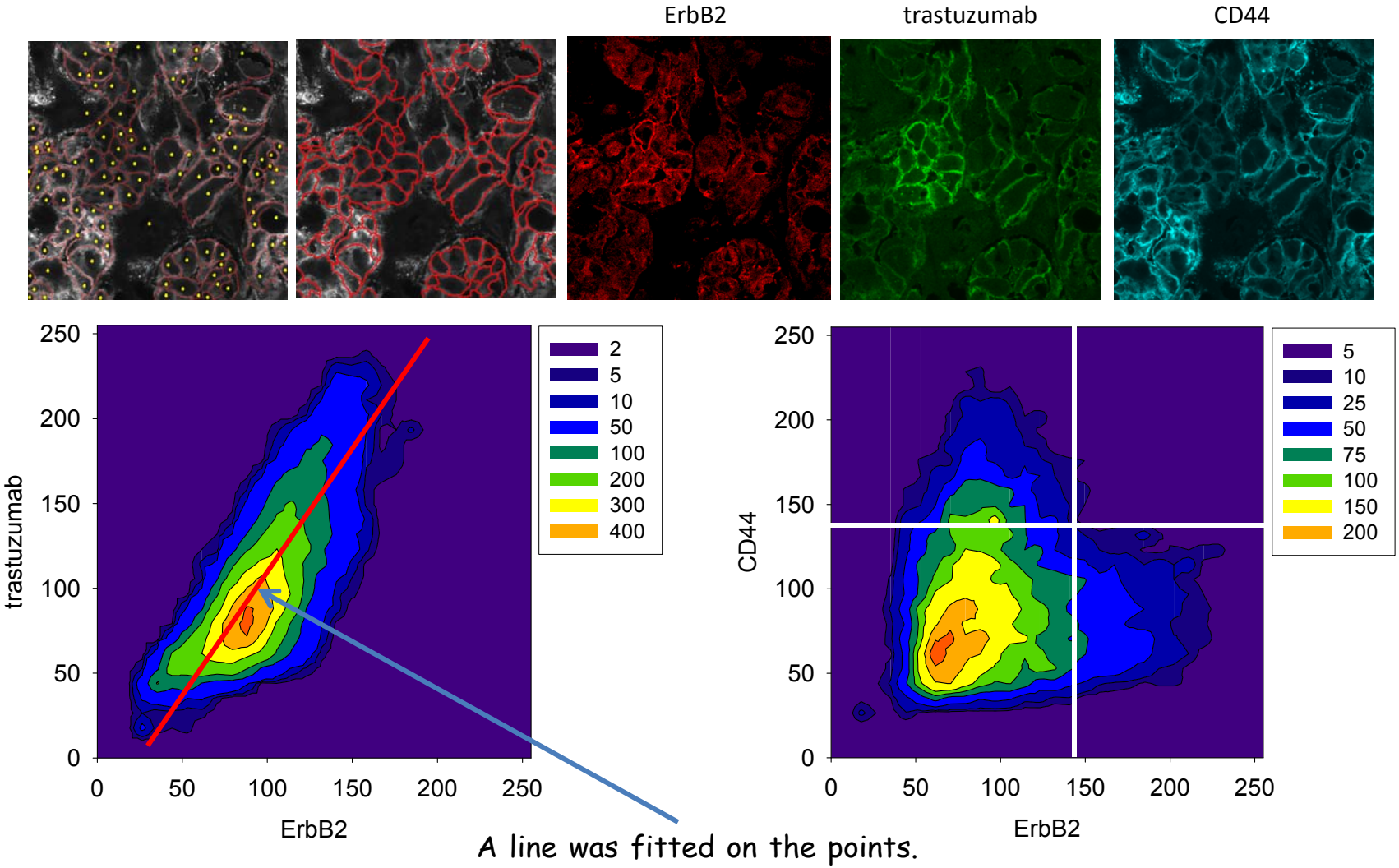
A statistic introduced by Karl Pearson (1857-1936) for the investigation of the relationship between two random variables (fluorescence intensities):

$$\rho = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sigma_X \sigma_Y} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-\left[E(X)\right]^2}\sqrt{E(Y^2)-\left[E(Y)\right]^2}}$$

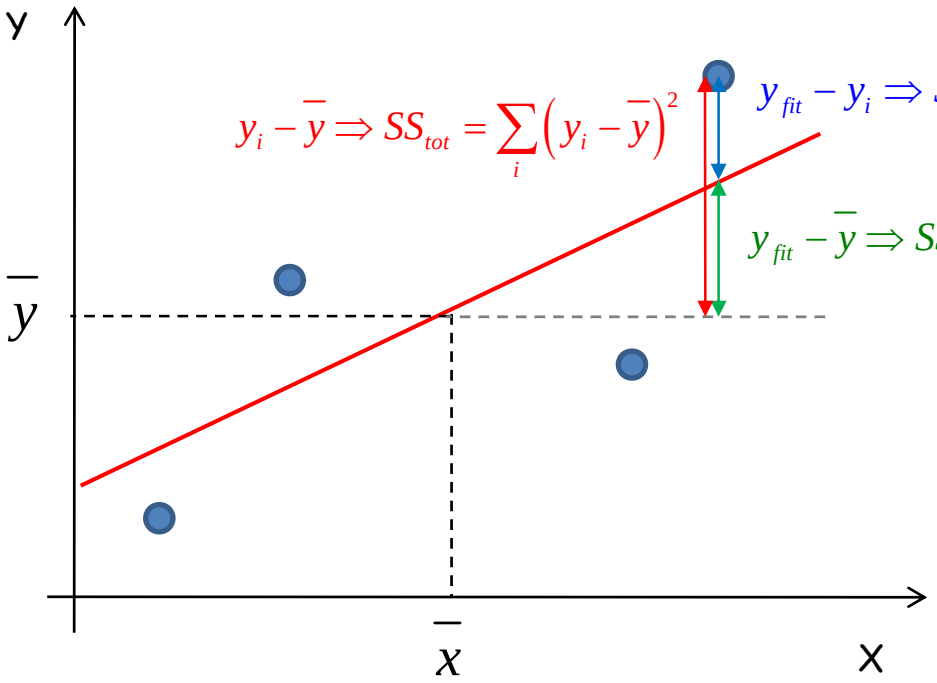
$$r = \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2}\sqrt{\sum_i (y_i-\bar{y})^2}} = \frac{n\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{\sqrt{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}\sqrt{n\sum_i y_i^2 - \left(\sum_i y_i\right)^2}}$$

- It measures the **LINEAR** relationship between the variables, i.e. how well a **STRAIGHT LINE** can be fitted to the x-y points.
- Its range is between -1 and 1:
  - 1 – perfect linear correlation
  - 0 – complete absence of linear correlation
  - -1 – perfect anticorrelation

Displaying pixel intensities: dot plot, density plot



Properties and problems of the Pearson correlation coefficient 1.



$$SS_{reg} + SS_{err} = SS_{tot}$$

Goodness of fit (coefficient of determination):

$$\frac{SS_{reg}}{SS_{tot}} = 1 - \frac{SS_{err}}{SS_{tot}} = r^2$$

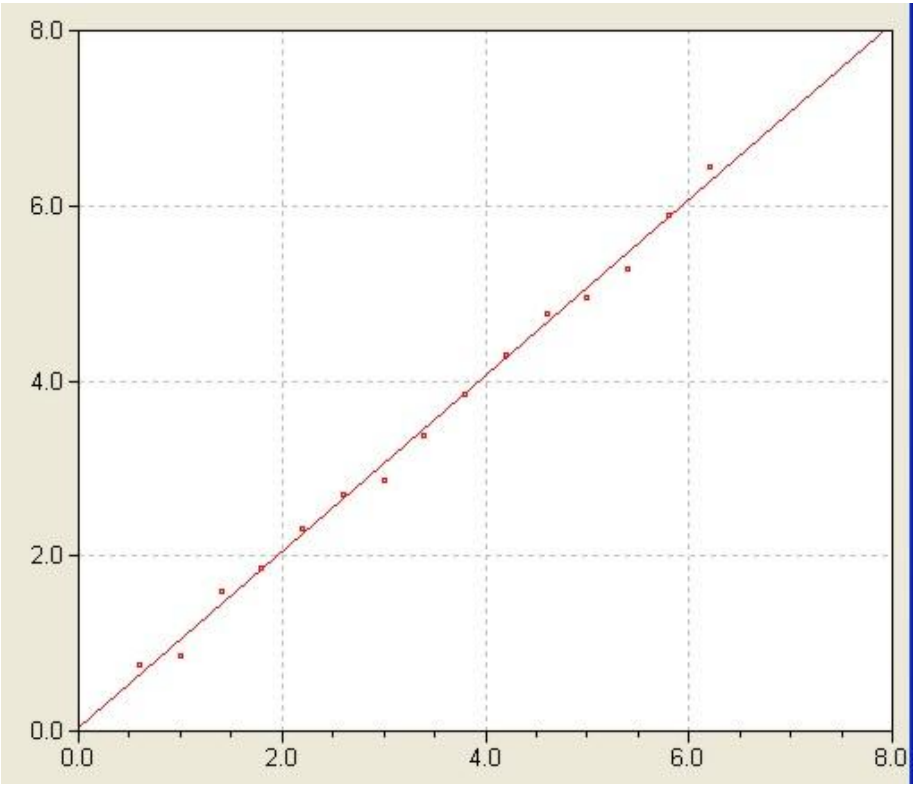
Goodness of fit estimates how many % of the variance of the dependent variable (y) can be accounted for by the variability of the dependent variable (x) e.g.:

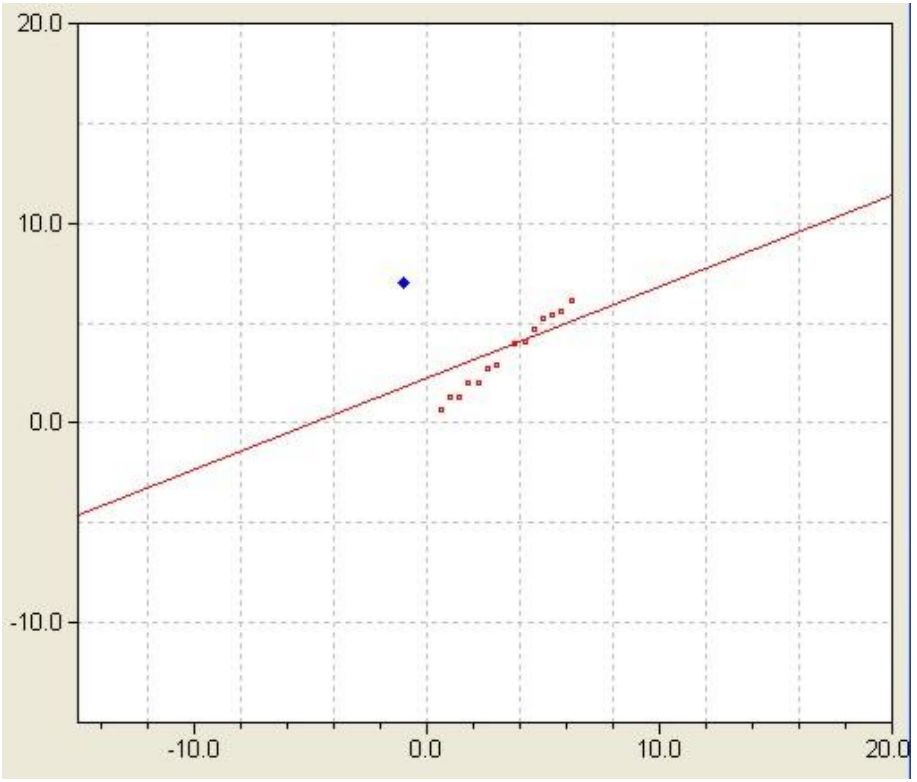
correlation coefficient, r=0.7

$r^2 = 0.49$

In the case of this relatively large correlation coefficient approximately 50% of the variability of the dependent variable is due to the error term (SS_{err}).

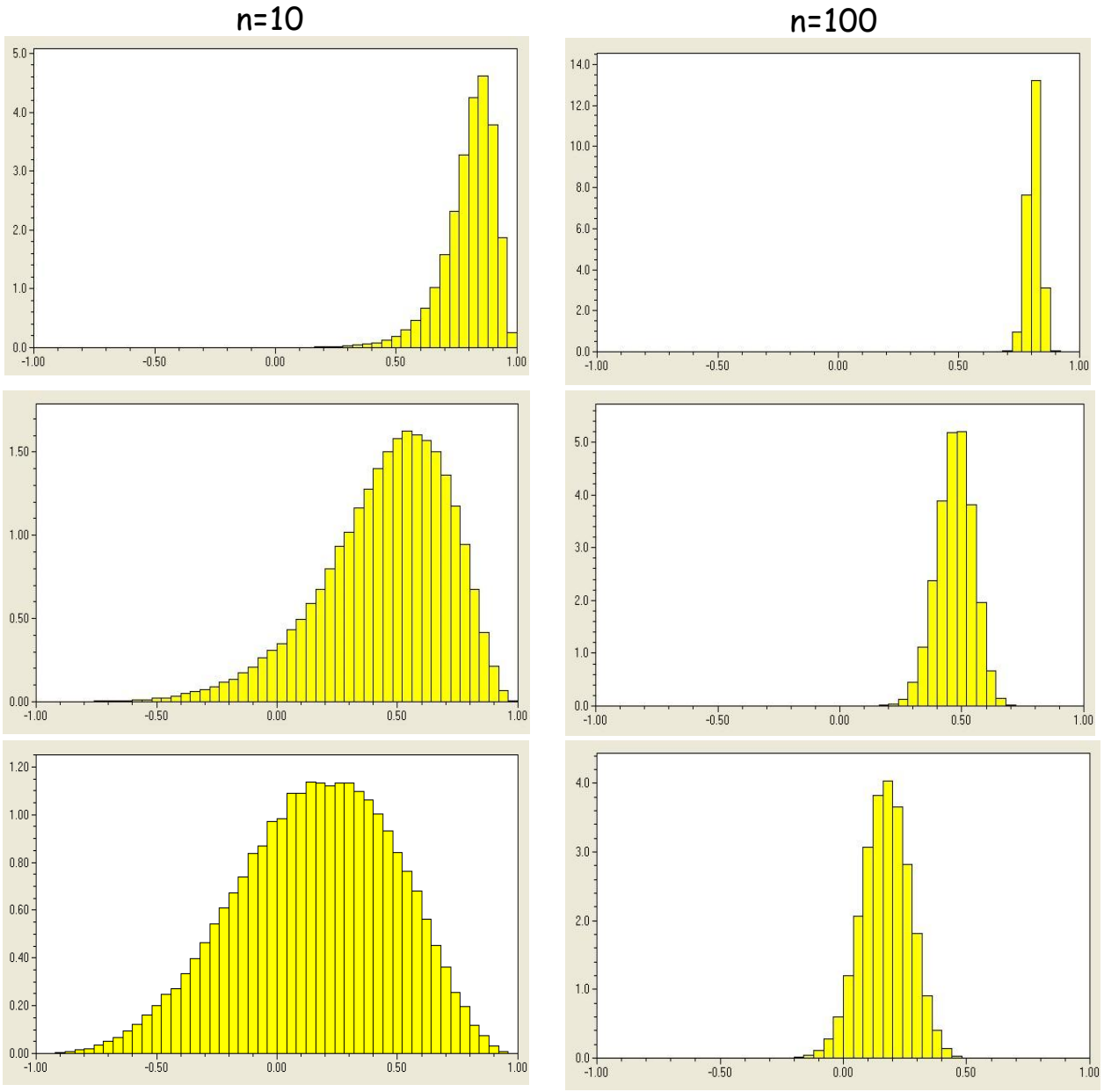
Properties and problems of the Pearson correlation coefficient 2.





A single outlier value significantly deteriorates the correlation.

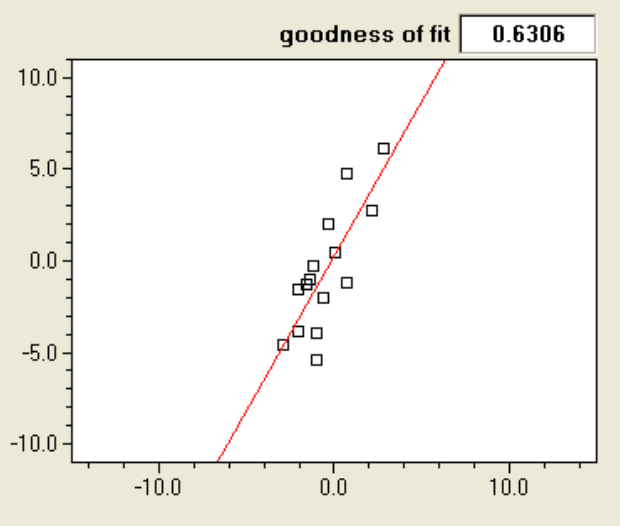
Properties and problems of the Pearson correlation coefficient 3.

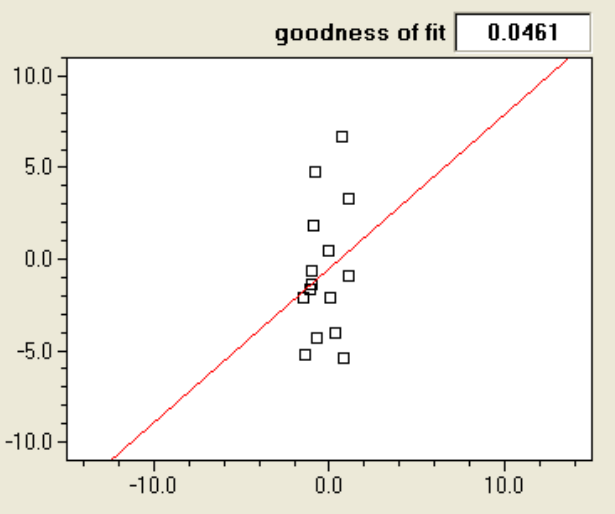


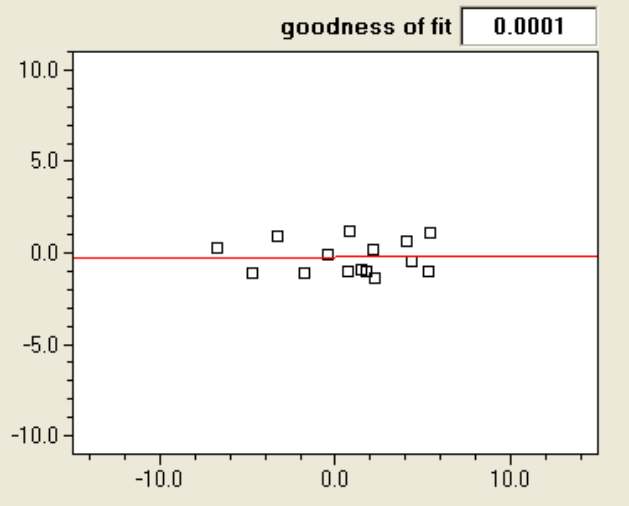
- The S.D. of the correlation coefficient is large if the number of elements is low.
- The smaller the correlation coefficient, the larger its S.D. is.

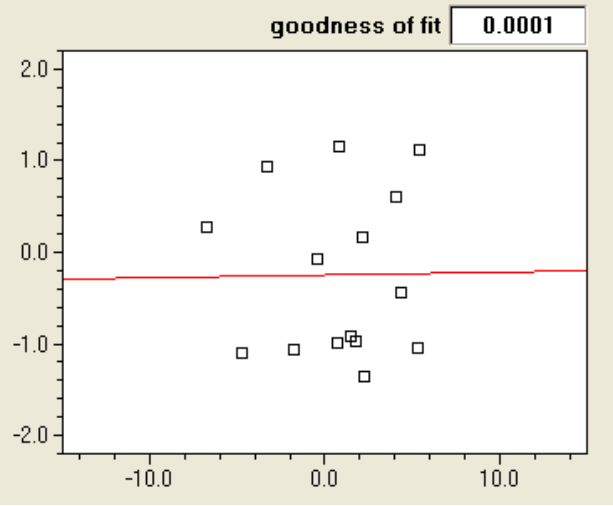
A small correlation coefficient obtained with low number of points is unreliable.

Properties and problems of the Pearson correlation coefficient 4.









If $SD_x \ll SD_y$, then the correlation depends strongly on small variations of x.

Determining the correlation coefficient only makes sense if the variabilities of the x and y variables are comparable.

If $SD_y \ll SD_x$, then the goodness of fit is bad.

After zooming on the y-axis it can be seen that only a small fraction of the variability in y is explained by the fit.

Properties and problems of the Pearson correlation coefficient 5.

$$slope = r\frac{SD_y}{SD_x} \Rightarrow r = slope\frac{SD_x}{SD_y}$$

If the slope is too low, the correlation coefficient will also be low even if the fit is reasonably good (assuming that SD_x/SD_y is not very large) .

Statistical tests for the correlation coefficient

1. t-test

$$t_{n-2} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Null hypothesis: ρ=0 (there is no linear correlation)
It can only be used to test whether the correlation coefficient is zero.

2. Fisher's z-transform

The calculated (r) and the assumed (ρ) correlation coefficients have to be transformed to follow a normal distribution:

$$z(r) = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right),\ z(\rho) = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

The statistic to be calculated for the *z* values:

$$z = \frac{z(r) - z(\rho)}{\sqrt{\frac{1}{n-3}}}$$

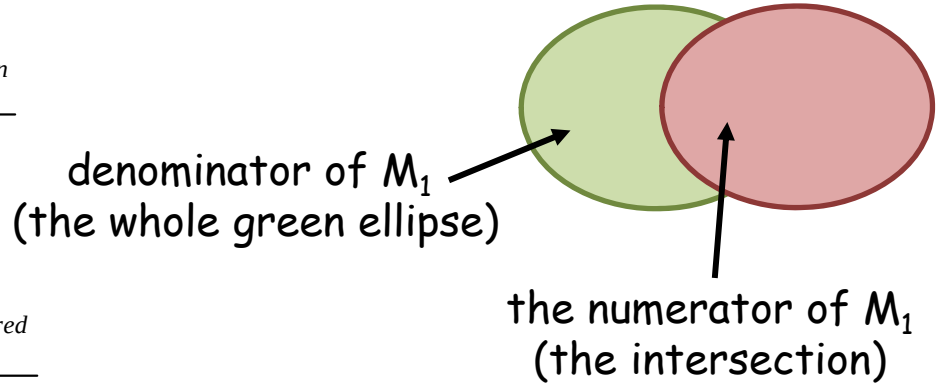
which follows a standard normal distribution.

Null hypothesis: r = ρ or r $\leq \rho$

ICCB 2: Manders coefficient

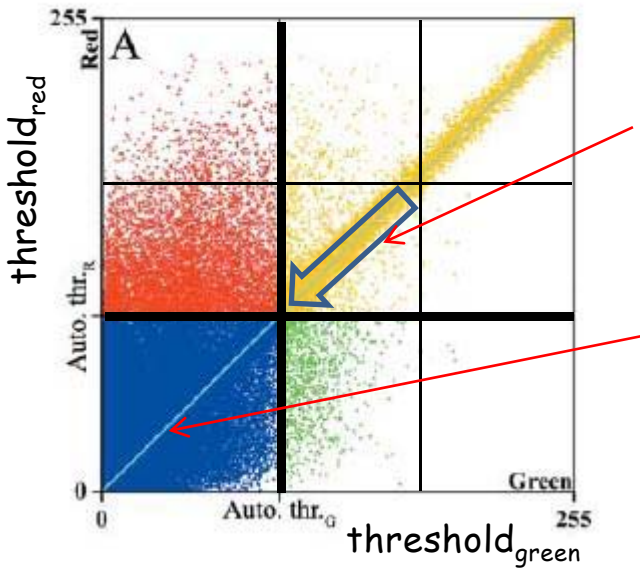
$$M_1 = \frac{\sum_{\text{above the threshold in the red channel}} I_{green}}{\sum_{\text{all pixels}} I_{green}}$$

$$M_2 = \frac{\sum_{\text{above the threshold in the green channel}} I_{red}}{\sum_{\text{all pixels}} I_{red}}$$



- its range is between 0 and 1
- sensitive to noise in the images

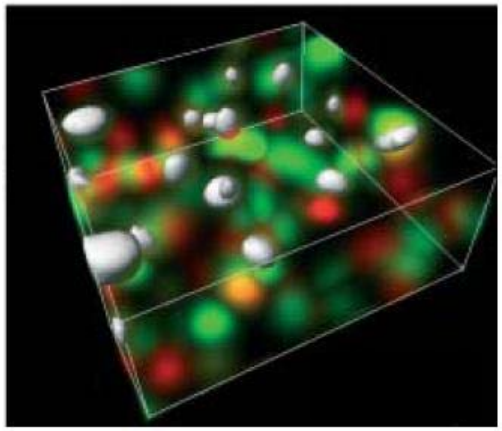
ICCB 3: Costes' method

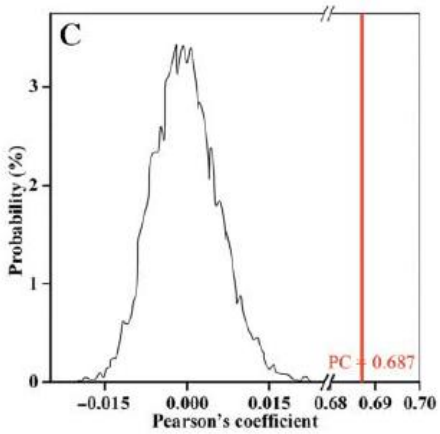


- The threshold is decreased until the correlation coefficient calculated for the under-threshold values (blue area) is zero.
- The yellow area corresponds to pixels exhibiting cololcalization.

$$I_{RED} = a\ I_{GREEN} + b$$

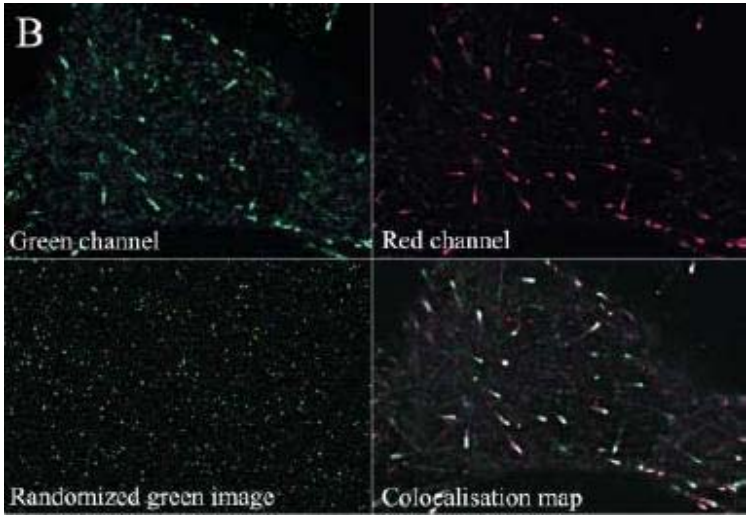
$$thr_{RED} = a\ thr_{GREEN} + b$$



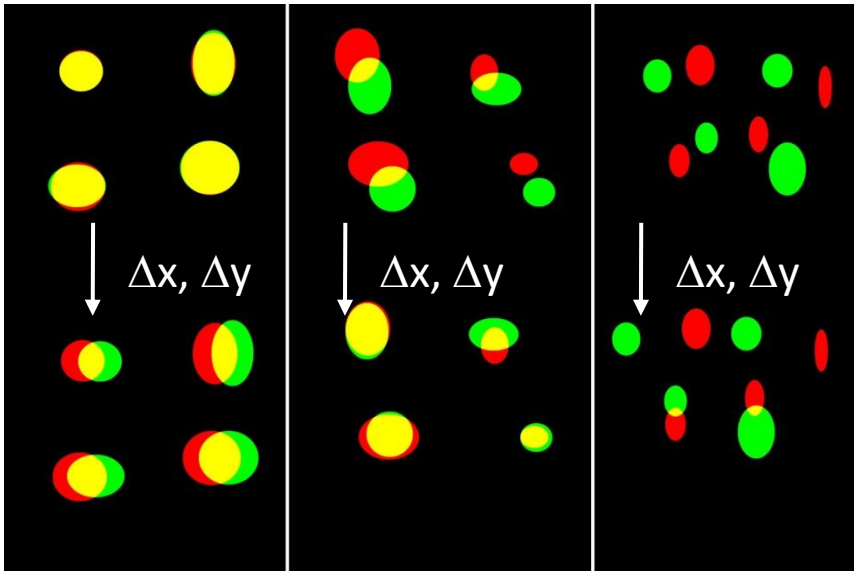


Determining significance of correlation:

- The pixels of one of the images are randomly reshuffled and the correlation coefficient between this scrambled image and the other image is determined.
- The above procedure is repeated several hundred-times yielding the distribution of *r* for random images.
- If the correlation coefficient for the original images is outside the 95% confidence interval, the correlation is significant.



ICCB 4: van Steensel's method

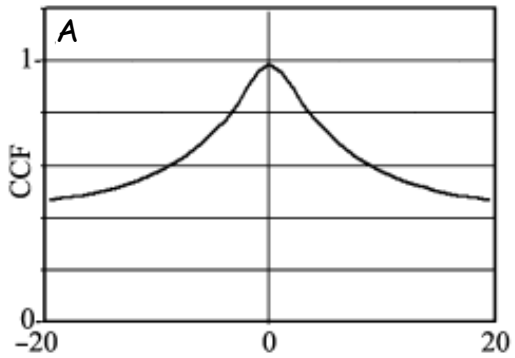


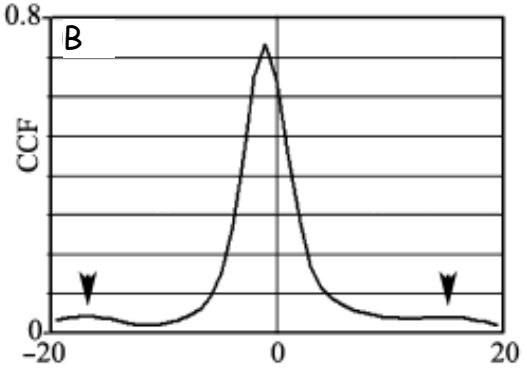
A. complete colocalization

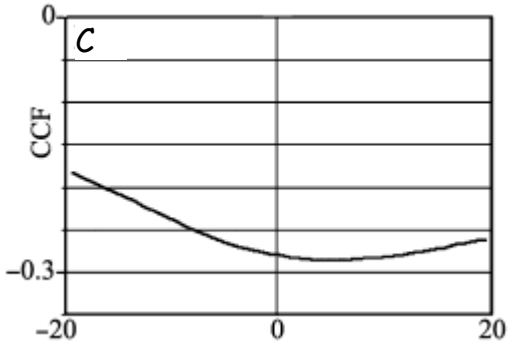
B. partial colocalization

C. exclusion

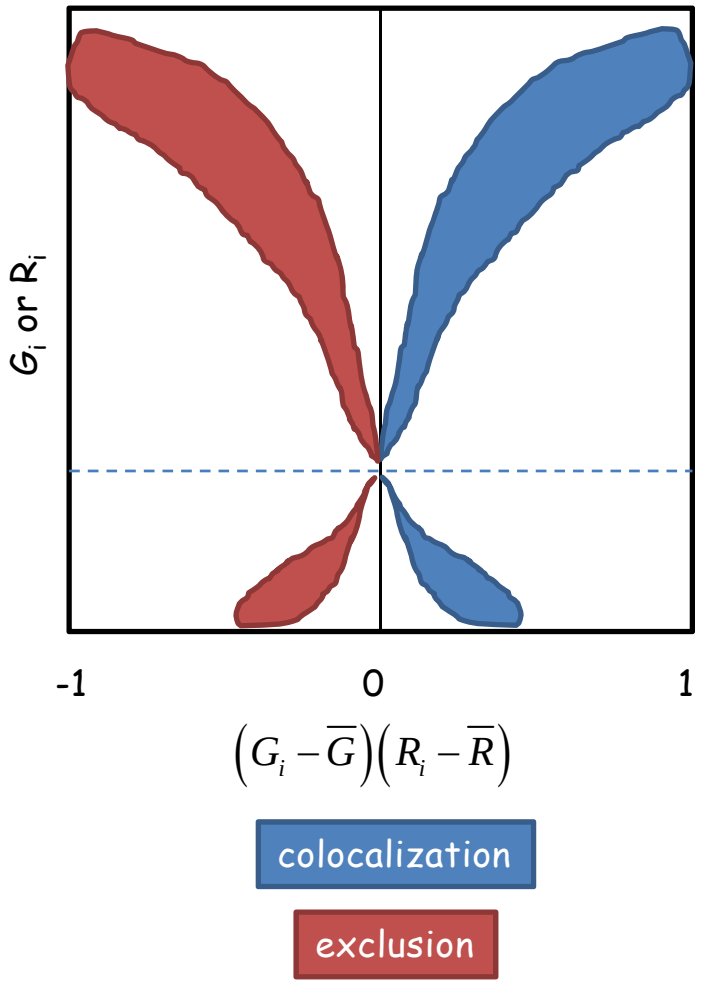
- The green image is shifted in *x* and *y* directions relative to the red one.
- The correlation coefficient (CCR – cross-correlation coefficient) is determined after each step.
- In the case of colocalization correlation (the area of the yellow regions) is expected to decrease.







ICCB 5: Li's method



$$\sum_i \left(G_i - \overline{G}\right) = 0$$

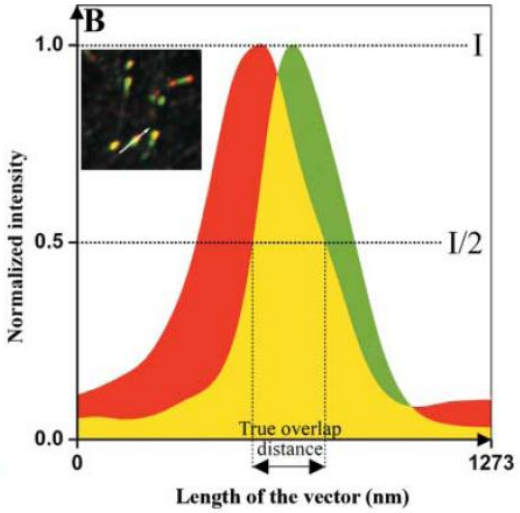
$$\sum_i \left(R_i - \overline{R}\right) = 0$$

If G_i deviates from $\overline{G}$ where R_i is different from $\overline{R}$ (correlation), then the value of $\sum_i \left(G_i - \overline{G}\right)\left(R_i - \overline{R}\right)$ will be positive.

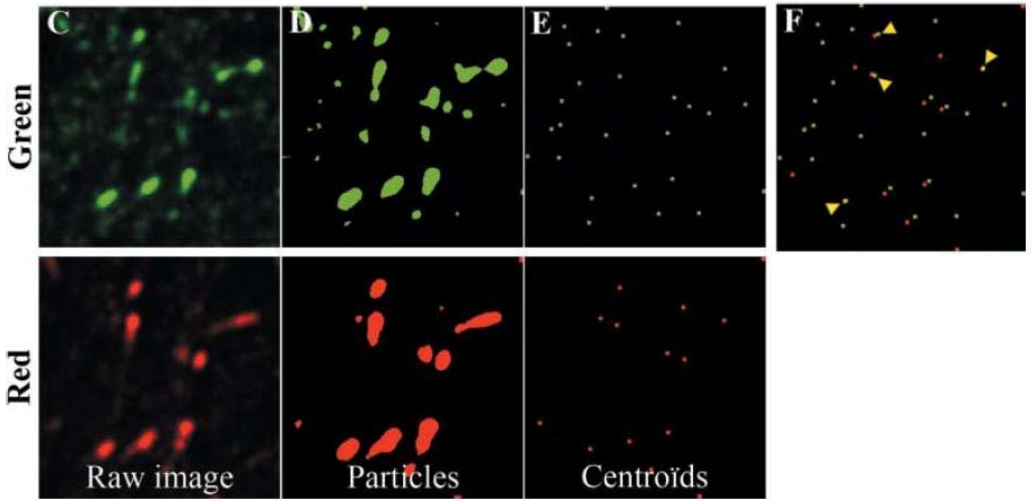
- It provides an easy-to-interpret graphical representation of colocalization.
- Quantitative evaluation: ICQ – intensity correlation quotient (the fraction of pixels in the positive region of the horizontal axis).

Object-based colocalization analysis

- Segmentation is always the first step of such an approach, i.e. the separation of foreground from background. Each pixel is assigned to either the background or the foreground (objects).
- After segmentation different parameters of the objects are analyzed:



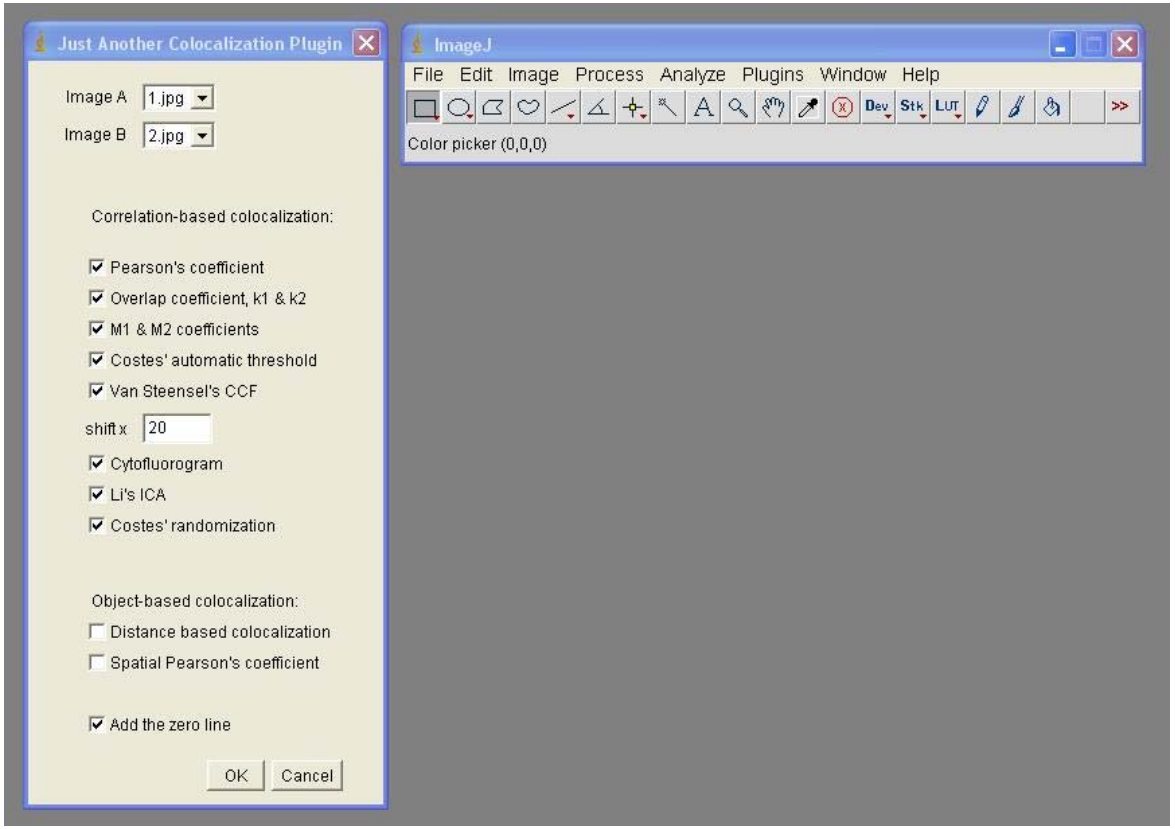
- The normalized intensity profile of the objects along a line is determined.
- If the length of that part where both curves are above $\frac{1}{2}$ („true overlap distance") exceeds the resolution limit of the microscope, colocalization is present.



- The centroids of objects in the red and green images are determined.
- If the centroids of the red and green objects are separated from each other with a distance less than the resolution limit of the microscope (yellow arrowheads in part F), then colocalization is present.

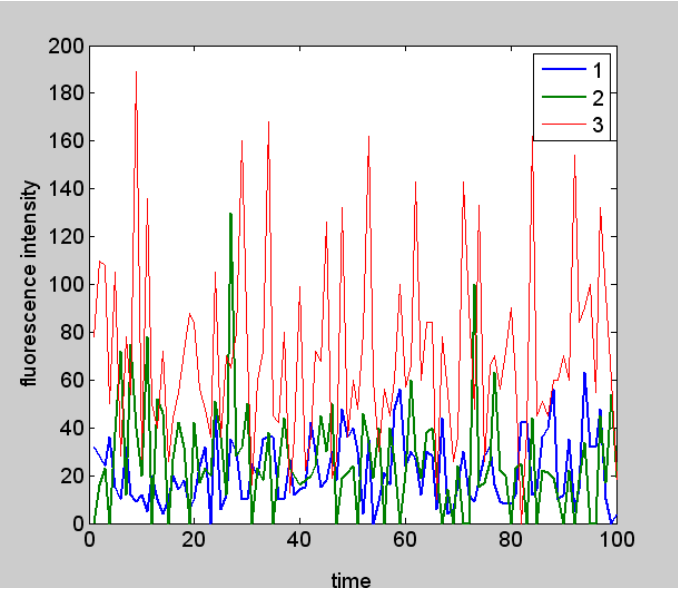
What kind of programs can be used to calculate colocalization?

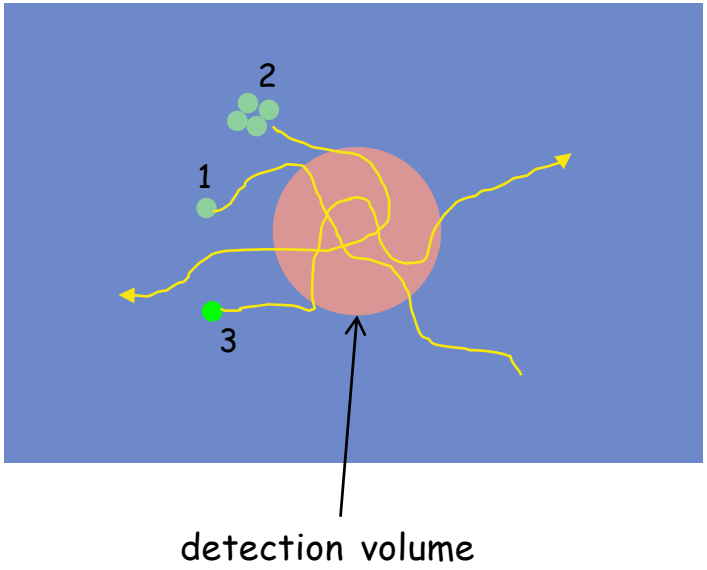
- Image J, JaCoP (Just Another Colocalization Plugin) and other plugins (http://rsb.info.nih.gov/ij)
- Matlab, Excel, etc…



FCS for the determination of molecular interactions 1.

- Principle of FCS (from the standpoint of molecular associations): fluorescence intensity fluctuations in a pixel (variance of fluorescence intensity) are related to the fluorescence intensity of diffusing units.

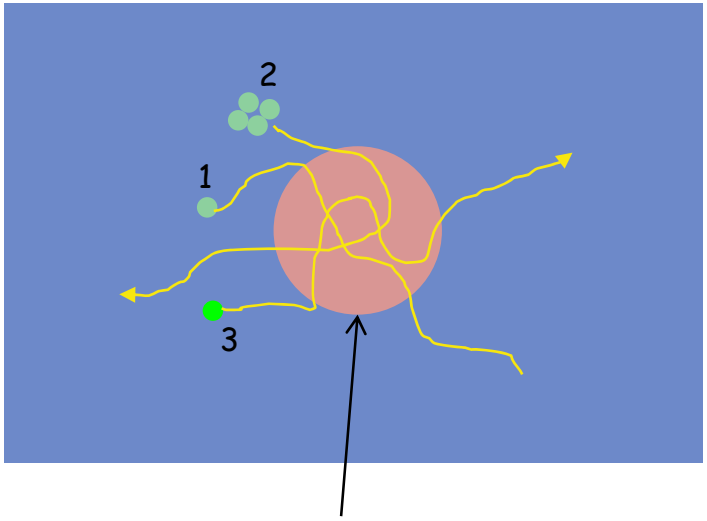




	mean	variance
1 – dim particle, c_1	21.9	204.6
2 – tetramer of dim particle, $c_1/4$	25.8	510.4
3 – 3x brighter particle, c_1	70.2	1520.6

FCS for the determination of molecular interactions 2.

- Pixel dwell time: the duration of time in which photons are collected from a single pixel
- Molecular brightness (ε): the number of photons detected from a single diffusing unit during the pixel dwell time.



$$\varepsilon_2 = 4\varepsilon_1$$

$$\varepsilon_3 = 3\varepsilon_1$$

For the determination of cluster size (monomer, dimer, etc.) the molecular brightness of a monomer has to be known.

determination of the distribution or variance of photon counts

↓

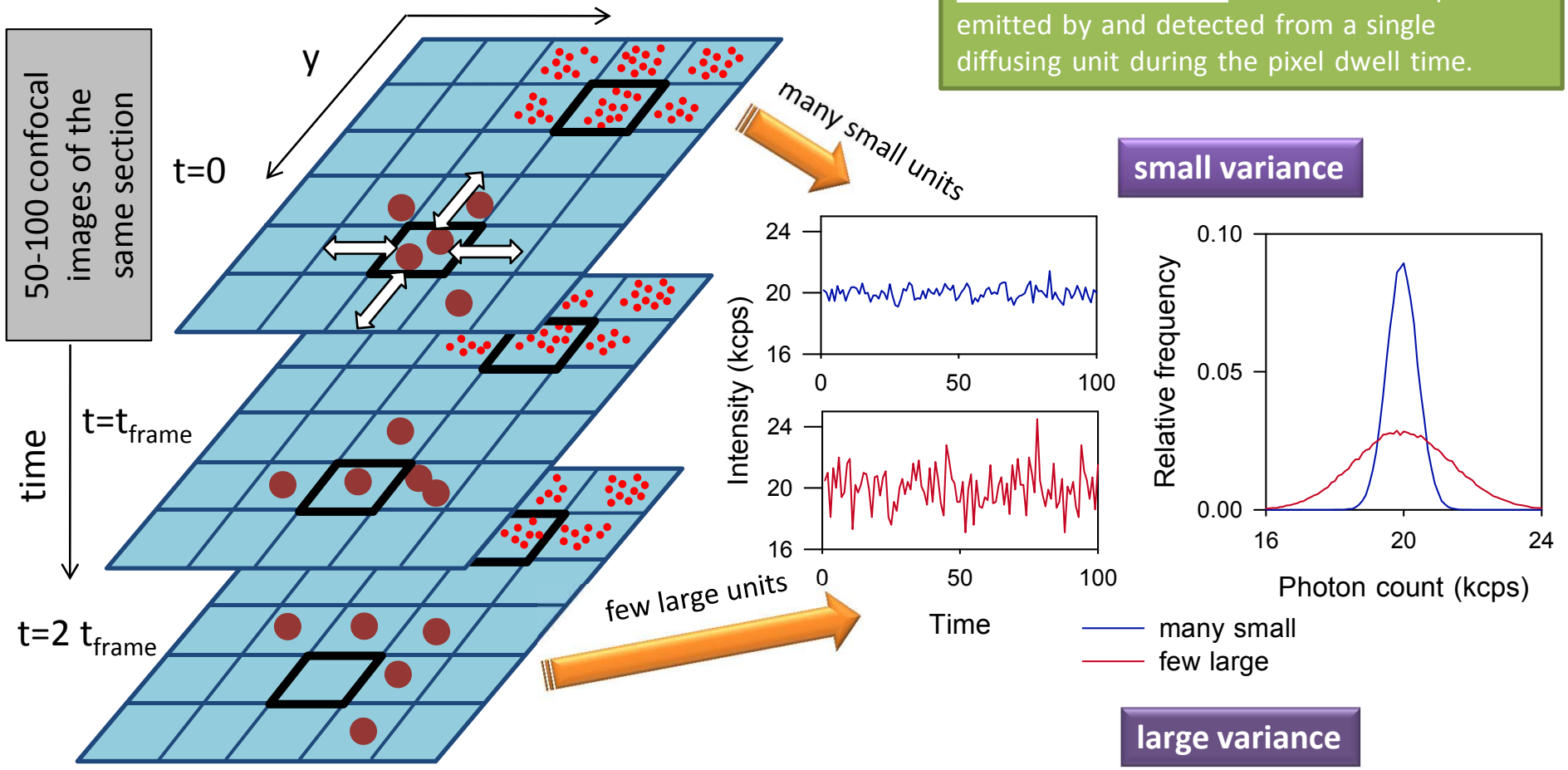
determination of molecular brightness

↓

determination of the state of association using the molecular brightness of a monomer

Number and brightness analysis (N&B)

Molecular brightness: the number of photons emitted by and detected from a single diffusing unit during the pixel dwell time.



Digman, M. A., R. Dalal, A. F. Horwitz, E. Gratton. (2008) Biophys J 94:2320-2332.

N&B analysis: sources of variation

Variance due to fluctuations of the particle number:

$$\sigma_n^2 = \varepsilon^2 n$$

Variance due to fluctuation in the number of photons:

$$\sigma_d^2 = \varepsilon n$$

Background information:

$$\sigma(k \cdot x) = k \cdot \sigma(x) \Rightarrow \sigma^2(k \cdot x) = k^2 \cdot \sigma^2(x)$$

n – average number of molecules in a pixel

n – average number of molecules

The number of molecules follows a Poisson distribution whose mean and variance are ***n***. We are, however, detecting photons.

$$\sigma^2(\text{photon number}) = \sigma^2(\varepsilon \cdot N) =$$
$$= \varepsilon^2 \cdot \sigma^2(N) = \varepsilon^2 \cdot n$$

N - momentary number of molecules

nε – average photon number

Since the photon number exhibits Poisson distribution (whose mean and variance are equal), the variance due to this factor is ***nε***

variance due to fluctuation in occupation number

$$B = \frac{\sigma^2}{\langle k \rangle} = \frac{\sigma_n^2}{\langle k \rangle} + \frac{\sigma_d^2}{\langle k \rangle} = \frac{\varepsilon^2 n}{\varepsilon n} + \frac{\varepsilon n}{\varepsilon n} = \varepsilon + 1$$

$$N = \frac{\langle k \rangle^2}{\sigma^2} = \frac{\varepsilon n}{\varepsilon + 1}$$

photon statistics (shot noise)

σ – pixel variance
⟨k⟩ – mean pixel intensity
n – expected value of occupation number
B – apparent brightness (count/dwell time/mol)
N – apparent number of molecules
ε – molecular brightness

If σ_n=0 (no fluctuation in the number of molecules, immobile molecules):

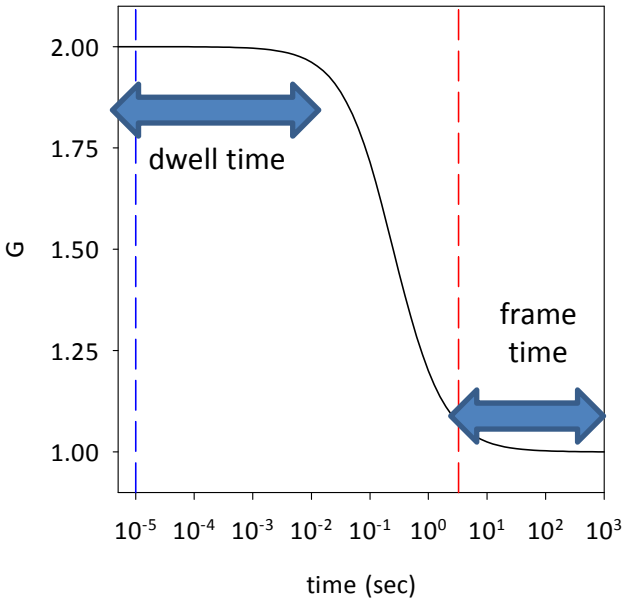
$$B = \frac{\sigma^2}{\langle k \rangle} = \frac{\sigma_d^2}{\langle k \rangle} = \frac{\varepsilon n}{\varepsilon n} = 1$$

- The apparent brightness of immobile molecules is 1 (independent of the molecular brightness).
- The requirement of the N&B method is that the molecules be mobile (otherwise B=1).

N&B analysis: proper selection of pixel dwell time and frame time

Requirements for the

- pixel dwell time: no averaging
- frame time: consecutive samplings of the same pixel are independent (long enough time between them)



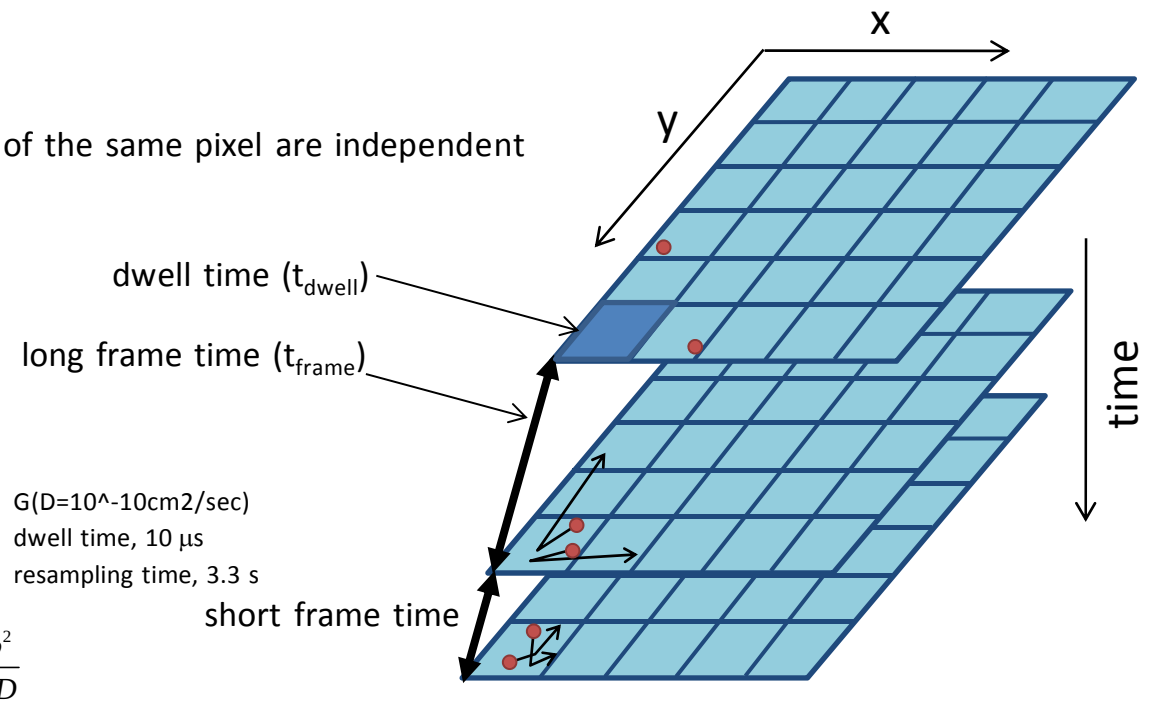
G(D=10^-10cm2/sec)
dwell time, 10 µs
resampling time, 3.3 s

$$\tau_D = \frac{\omega^2}{4D}$$

τ_D - diffusional correlation time

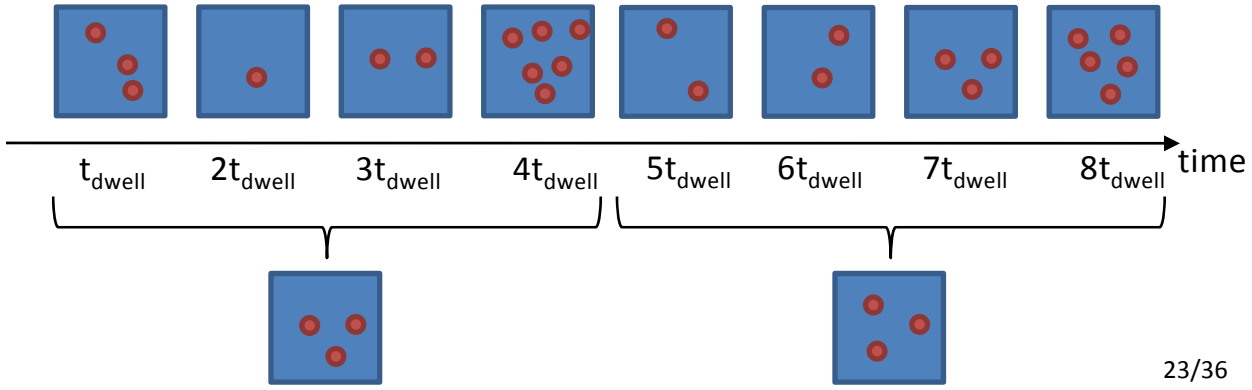
ω - radius of confocal detection volume

D - diffusion constant



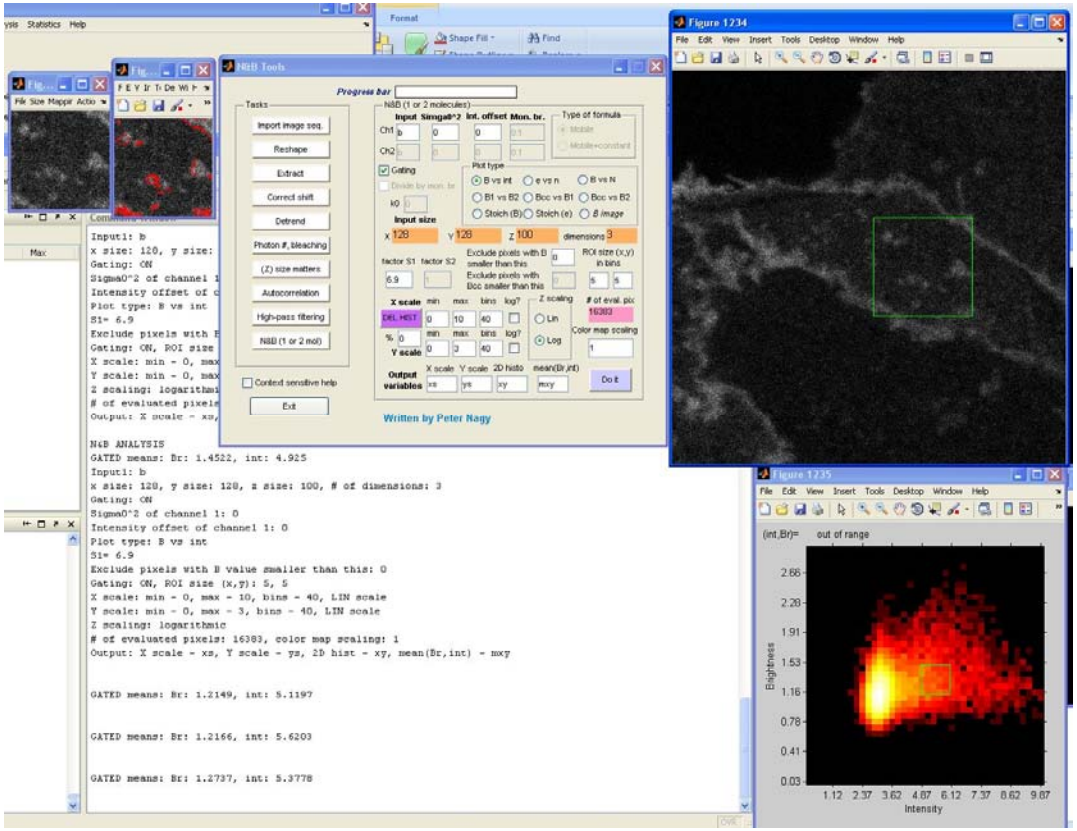
If the frame time is short, almost the same molecules are in the pixel in consecutive samplings → not independent, which is the requirement of Poisson statistics

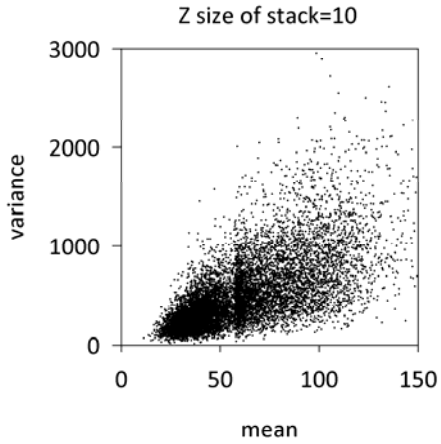
If pixels in the top row are detected for 4x longer time, the fluctuations in occupation number are averaged.

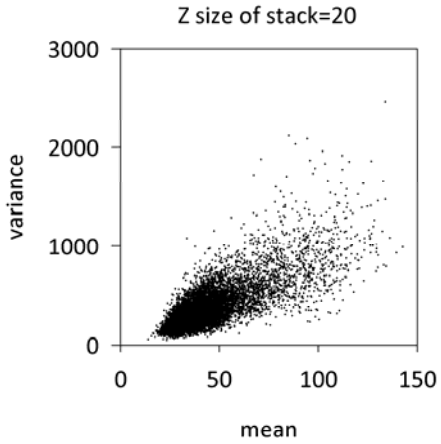


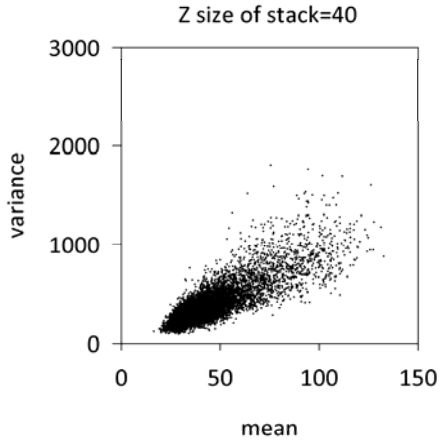
N&B measurements: evaluation

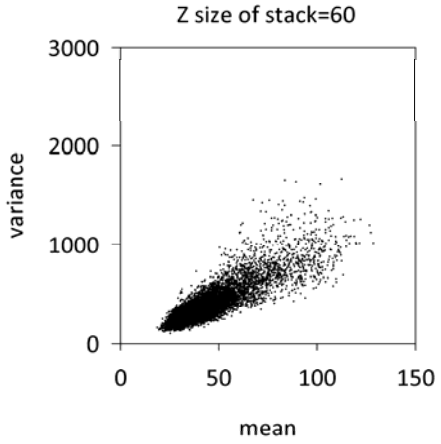
- Performing the measurement: a microscope equipped with a photon-counting detector is required. After calibration microscopes with analog detectors can also be used.
- Evaluation: special, dedicated software
  - Globals, a.k.a. SimFCS, http://www.lfd.uci.edu/globals/
  - Matlab, N&B Tools
- The method provides the molecular brightness and the number of molecules for every pixel.
- In the case of mixed molecule populations (e.g. 50% monomer, 50% dimer) the determination of the fraction of the different species is not always possible.

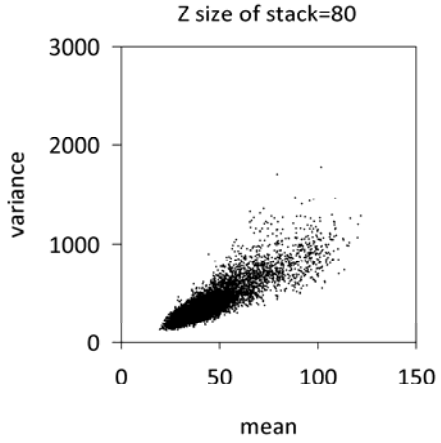


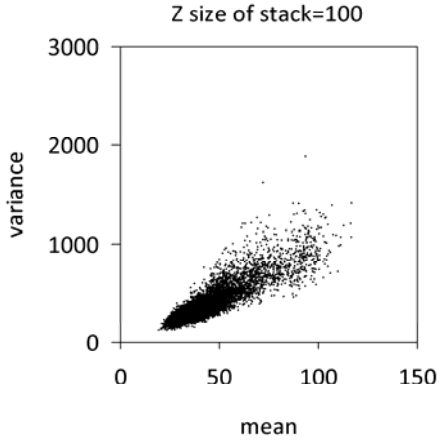












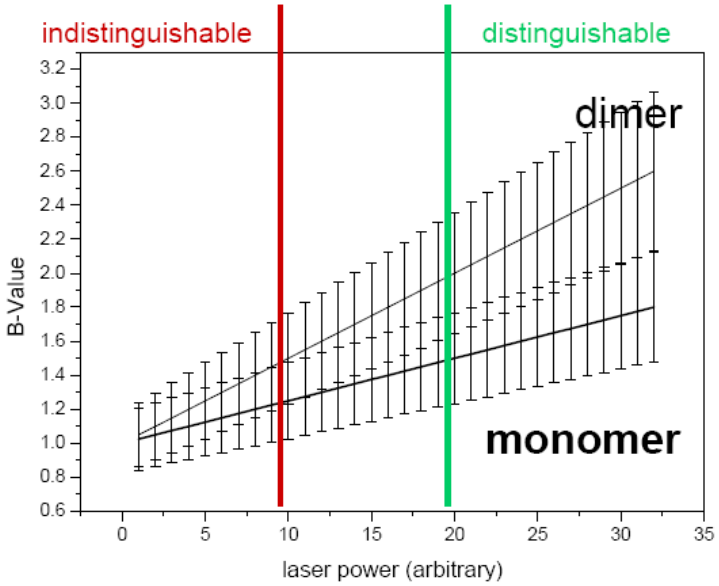
Problems in N&B measurements 1: number of frames

Variance has to be proportional to the mean.

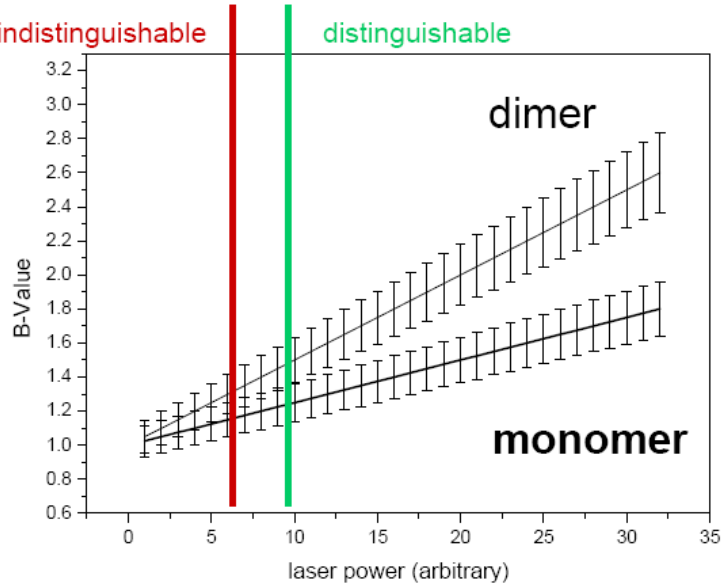
This is only achieved if the number of frames is high enough (due to measurement error).

Problems in N&B measurements 2: number of frames

B-Value and FWHM for 100 frames



B-Value and FWHM for 200 frames



Problems is N&B measurements 3: immobile molecules

If immobile molecules are present, their number of molecules/pixel will not change, i.e. their σ_n^2 contribution to the total variance is zero:

$$B^* = \frac{\sigma^2_{n,mobile} + \sigma^2_{d,mobile} + \sigma^2_{d,immobile}}{k_{mobile} + k_{immobile}} \quad \left(\sigma^2_{n,immobile} = 0\right)$$

$$B^* = \frac{\sigma^2_{n,mobile} + \sigma^2_{d,total}}{k_{total}} = \frac{f_M \sigma^2_{n,total} + \sigma^2_{d,total}}{k_{total}} \quad \left(\text{variance is additive, } f_M - \text{ fraction of mobile molecules}\right)$$

$$B^* = \frac{f_M n_{total} \varepsilon^2 + n_{total} \varepsilon}{n_{total} \varepsilon} = 1 + f_M \varepsilon$$

The measured apparent brightness (B^*) is decreased if immobile molecules are present.

$$B = 1 + \frac{B^* - 1}{f_M}$$

Problems is N&B measurements 4: photobleaching, stage shift

Photobleaching of immobile components and sample movement (stage shift):

they introduce another term into the variance → variance and brightness are increased.

Solution: high-pass filter the time series so that low frequency (slow) trends are eliminated.

Photobleaching of mobile components:

usually assumed to be negligible due to the replacement of bleached mobile molecules by unbleached ones.

If replacement doesn't take place, variance and brightness are underestimated.

N&B: determination of factor *S* (intensity unit/photon)

- Even some photon counting microscopes don't provide true photon counts, but an intensity unit proportional to the photon count (pseudo photon counting detectors).
- The intensity units have to be calibrated in terms of photon numbers:

$$number\ of\ photons = \frac{intensity\ unit}{S}$$

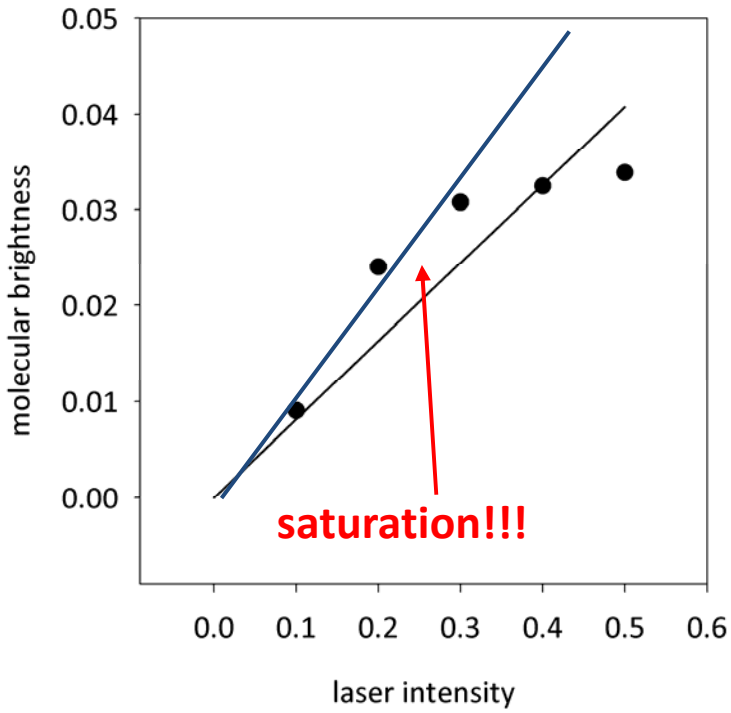
Using immobile particles:
Since the apparent brightness (*B*) of immobile particles is unity, such an *S* factor has to be chosen with which this is achieved.

By varying laser intensity:
Molecular brightness (ε=*B-1*) has to be proportional to laser intensity:
$\varepsilon \sim$laser intensity (I_L)
$\varepsilon = a\ I_L$, where *a* is a constant.
This is true if the y-intersection of the line fitted on the ε vs. I_L plot is zero. Such an *S* factor has to be chosen with which this is achieved.

In the case of detector saturation assumptions made for the variance-mean intensity relationship are not valid.

⬇

N&B measurements can only be made with an unsaturated detector.

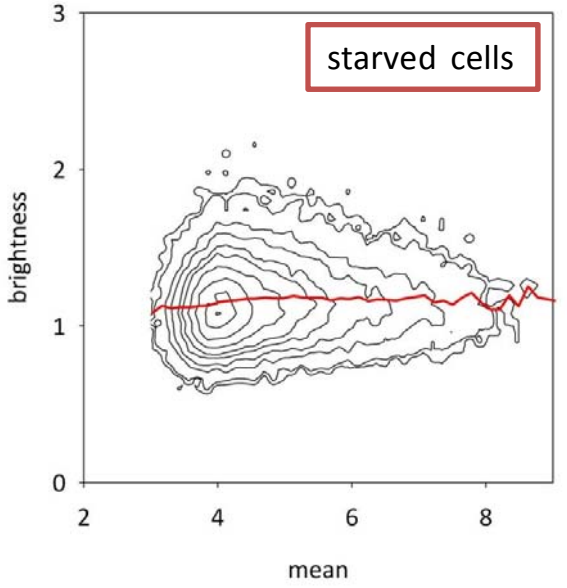


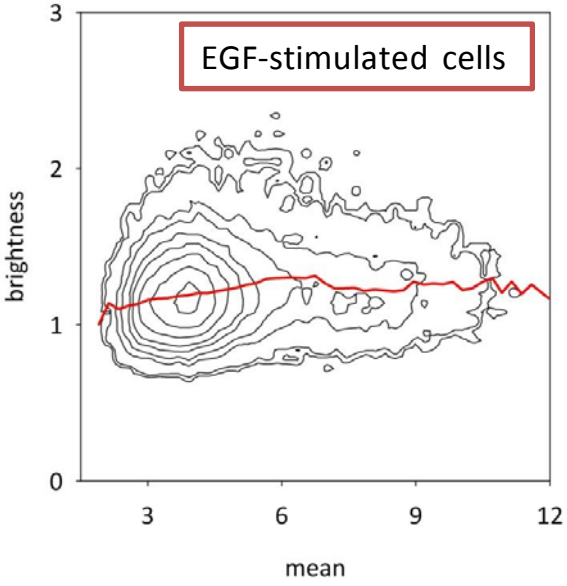
S=6.93 (by both methods,
fit on the first two points → blue line)

N&B analysis: what kind of *B* values do we expect?

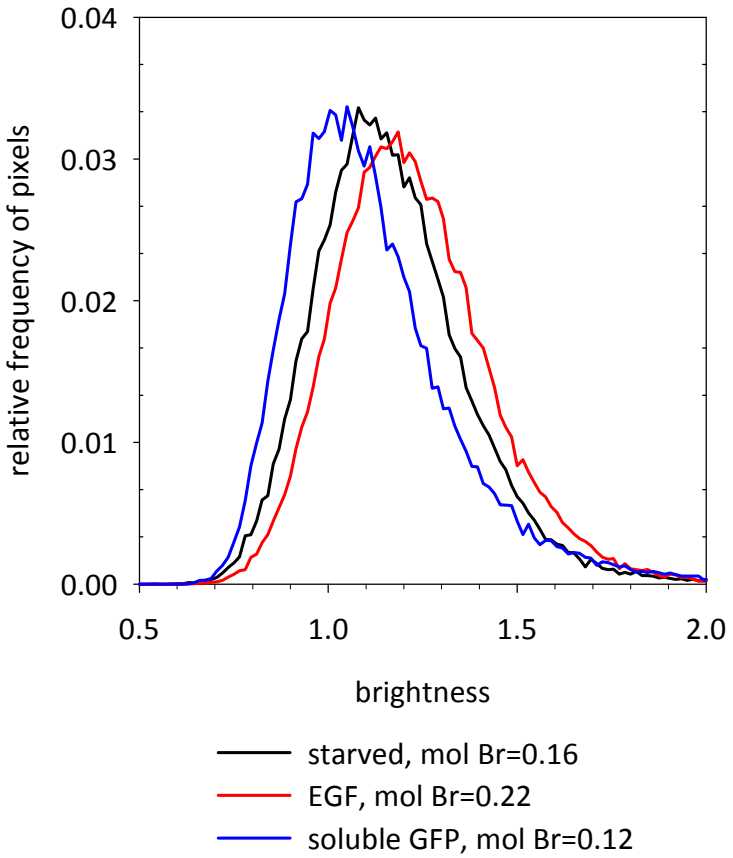
Adjustment	value of *B*		Remark
	eGFP (monomer) **eGFP (dimer)**	Alexa **Alexa (dimer)**	
max. laser power, long dwell time (50 µs)	1.25 **1.5**	6 **11**	fast photobleaching, phototoxicity
medium laser power, medium dwell time (32 µs)	1.08 **1.16**	2.6 **4.2**	good compromise
low laser power, short dwell time (4 µs)	1.002 **1.004**	1.04 **1.08**	*B* is too low

N&B measurements with cells expressing EGFR-eGFP

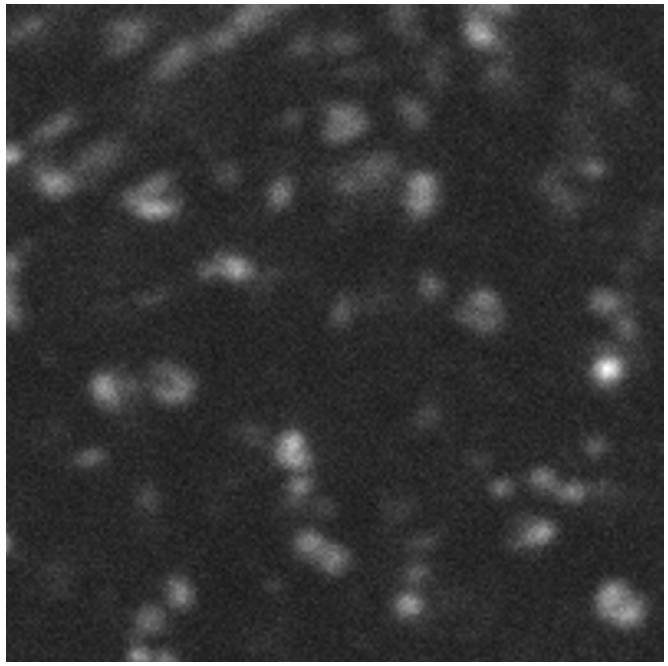




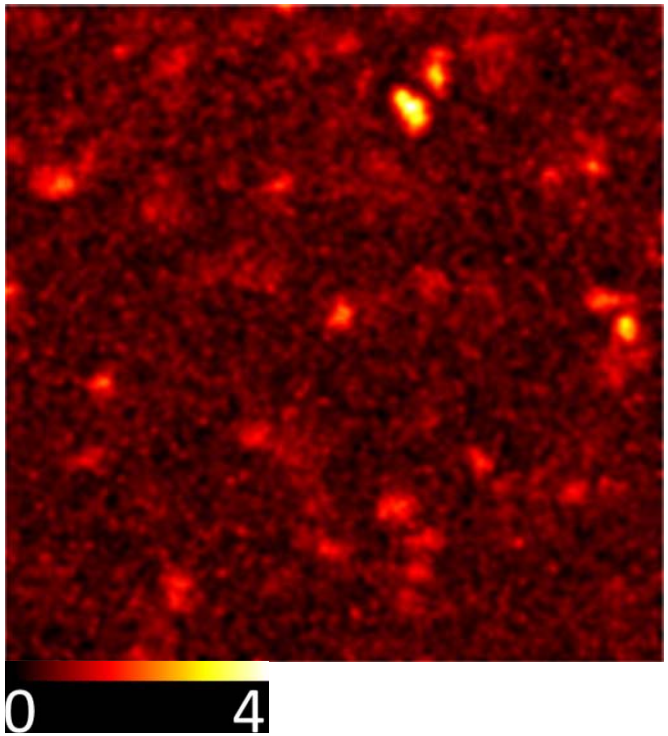
F1-4 cells: CHO cells stably transfected with ~6×10^5 EGFR-GFP/cell



N&B measurements with cells expressing EGFR-eGFP



Fluorescence intensity



Apparent brightness

N&B: pros and cons

Advantages:
- simple, interpretation of results is obvious
- it can be performed on practically any microscope (although photon counting microscopes are preferred)

Disadvantages:
- only mobile molecules can be measured
- all processes which modify variance (e.g. molecular blinking, photobleaching) interfere with the measurements
- mixed molecular populations complicate the measurements and the interpretation of results

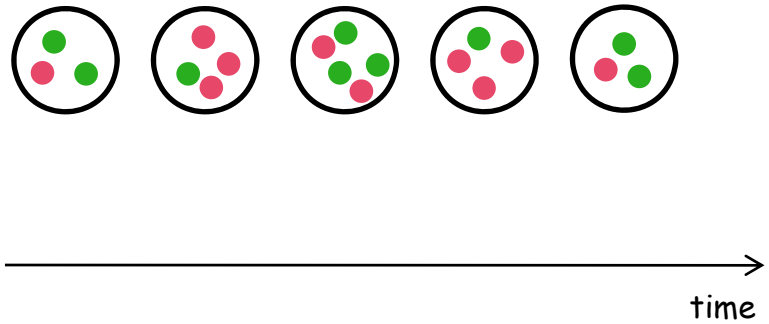
Good to consider:
- use low laser power (to avoid saturation and photobleaching)
- record many frames (to increase statistical precision/reproducibility)
- calibrate factor S (in the case of pseudo photon counting detectors)
- use appropriate „dwell time" and „repetition time" (frame time) settings

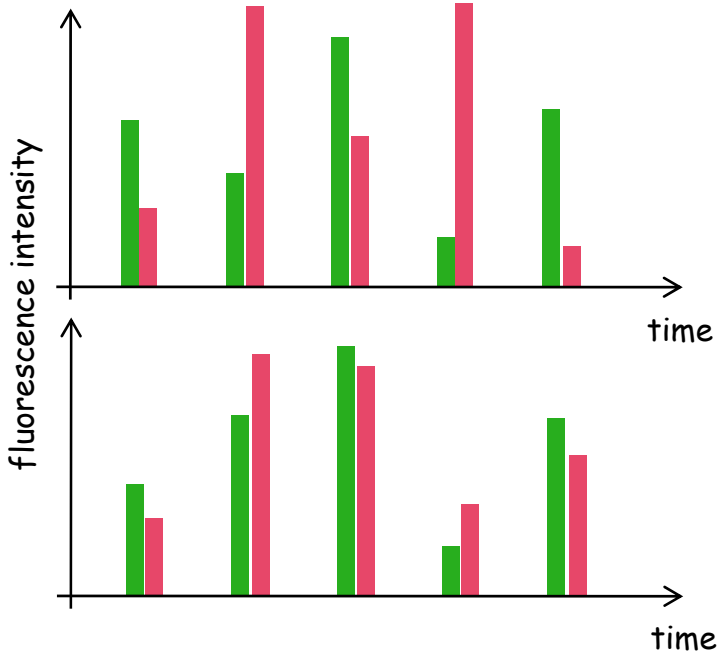
2-color N&B analysis (cross N&B): stoichiometry of molecular complexes

- Two different molecules are labeled by distinct fluorephores in a cell.
- Aim: determine whether the two molecules associate with each other, and if yes, what is the stoichiometry of the complex.
- Determine the molecular brightness (ε_1, ε_2) in both fluorescence changes using N&B analysis based on the variance in each channel.
- The values of ε_1 és ε_2 determined in this way do not prove that the molecules associate with each other since ε_1 and ε_2 were determined from the variance of the respective fluorescence channel which is independent of the variance of the other channel.
- Solution: Consider only those pixels in which the intensity fluctuations correlate with each other. This can be achieved by calculating the cross-variance.

Digman, M.A.,Wiseman P.W, Choi C, Horwitz A.R., Gratton E. (2009) PNAS 106: 2170-2175.

Cross-variance





If the two molecules are clustered, the two intensities correlate with each other → σ_{cc} will be larger than zero.

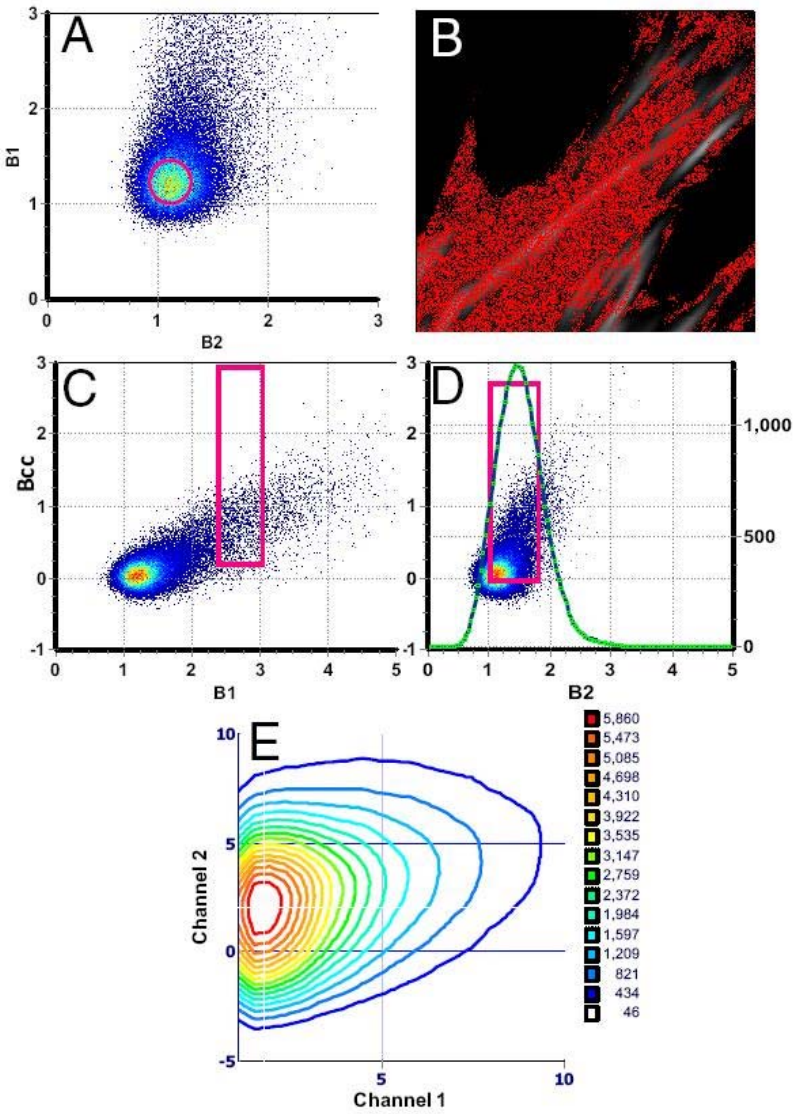
Cross-variance:

$$\sigma_{cc}^2 = \frac{\sum_{i=1}^{k}\left(G_i - \overline{G}\right)\left(R_i - \overline{R}\right)}{k}$$

Cross-brightness

$$B_{cc} = \frac{\sigma_{cc}^2}{\sqrt{M(G)M(R)}}$$

Determination of stoichiometry



- Based on the cross-variance we evaluate only those pixels in which the two molecules associate with each other.
- The stoichiometry can be determined using known molecular brightnesses.