Software for the nanoscale

Microscopy image analysis at the Cambridge Advanced Imaging Centre

Leila Mureșan

EPSRC Research Software Engineer Fellow
Cambridge Advanced Imaging Centre, University of Cambridge

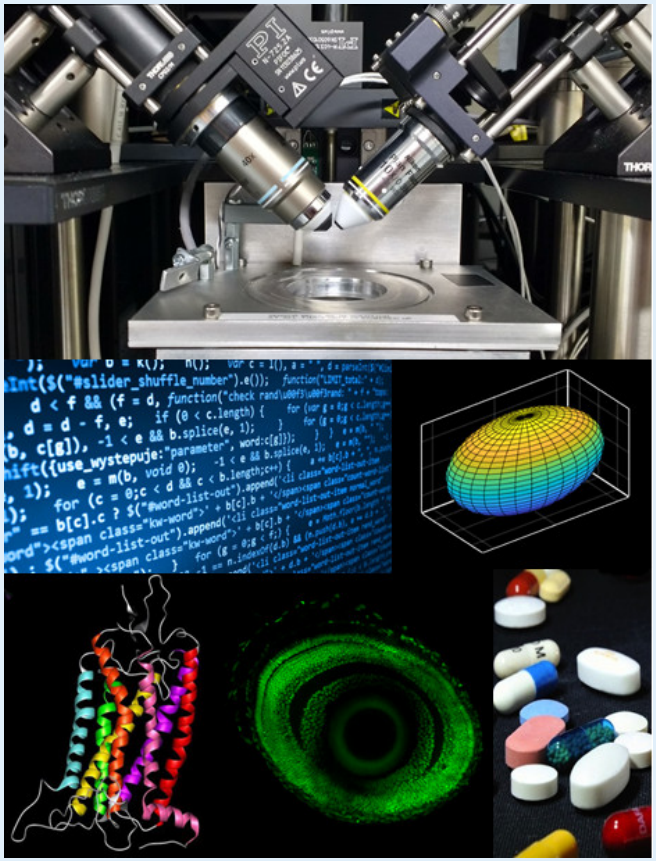
22 April 2020

Outline

Introduction

Cambridge Advanced Imaging Centre

- Research facility inaugurated in 2014.
- Located in the Dept. of Physiology, Development and Neuroscience, Anatomy building (School of Biology)
- **Aim:** develop and enable the exploitation of cutting edge optical systems



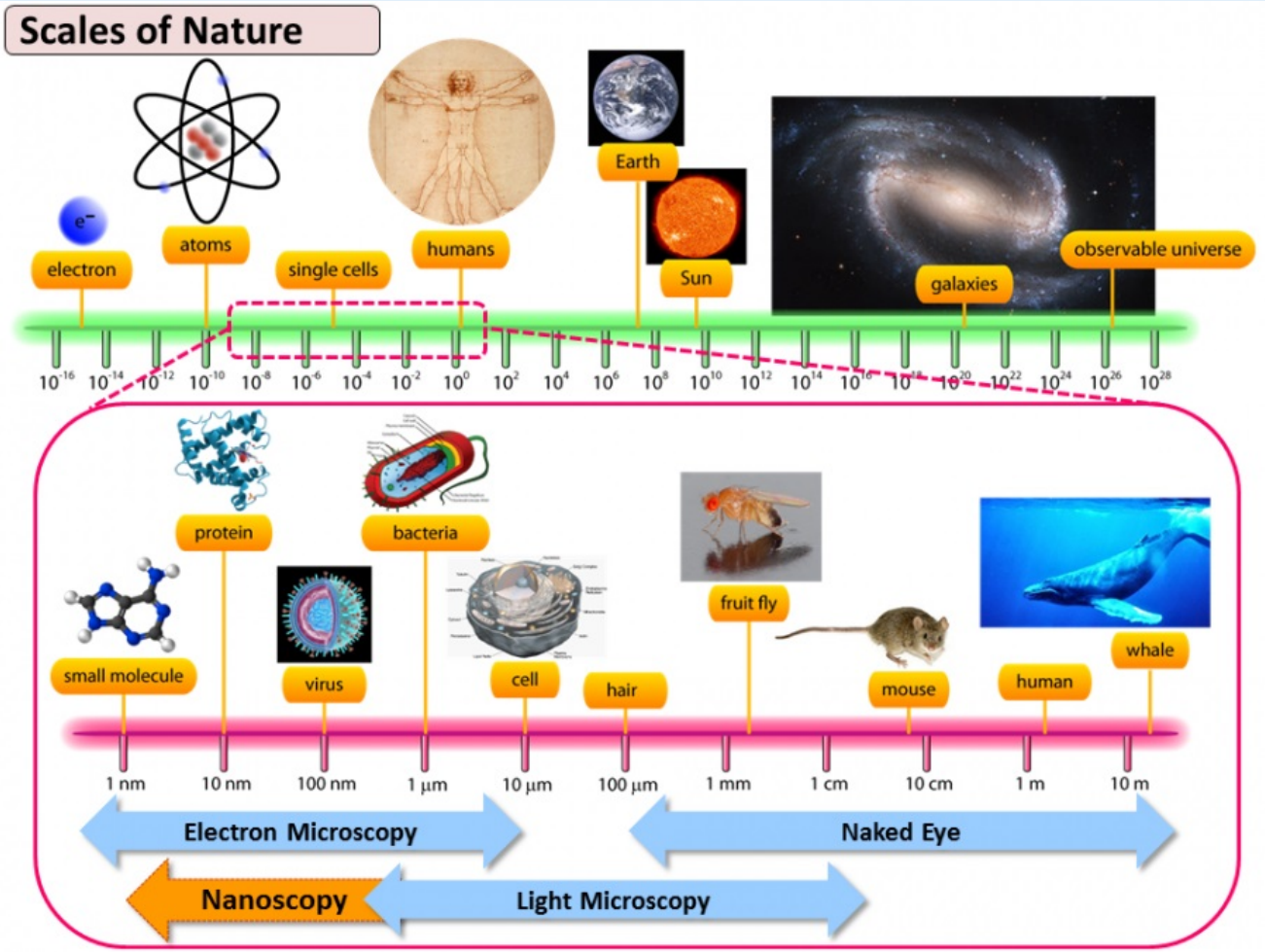
Cambridge Advanced Imaging Centre

Activity:

- Light sheet systems, super-resolution microscopes, several versatile hybrid systems
- 431 users from 161 groups (30 institutes and departments from UoC, other universities and institutes, 5 companies)

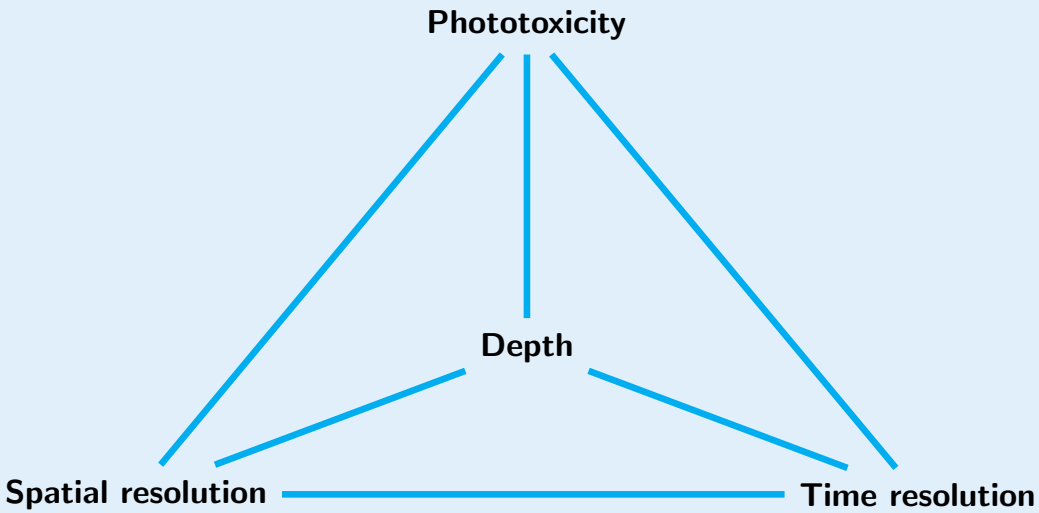
CAIC team

Resolution map



www.bates.edu/gould-research-lab/research/

Goals in fluorescence microscopy

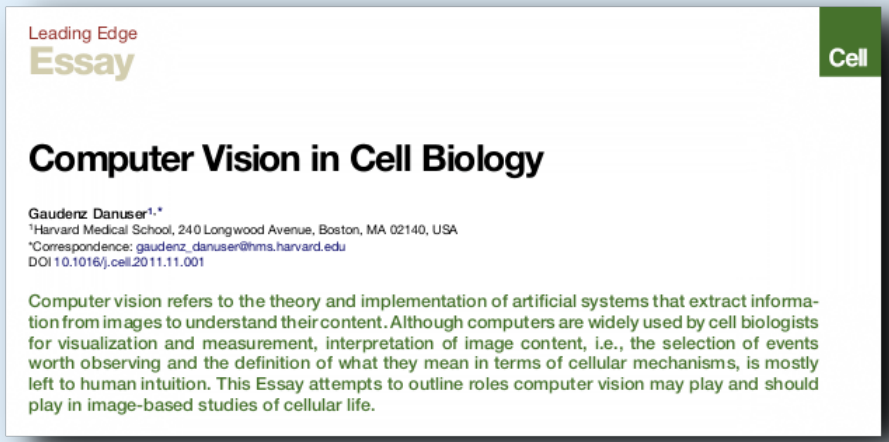


Computational microscopy has a pivotal role in achieving a better trade-off between goals

Bioimage informatics

Current limitations in bioimage-informatics techniques are preventing sophisticated optical methods from realizing their full potential.

The quest for quantitative microscopy, Nature Methods 9, 2012



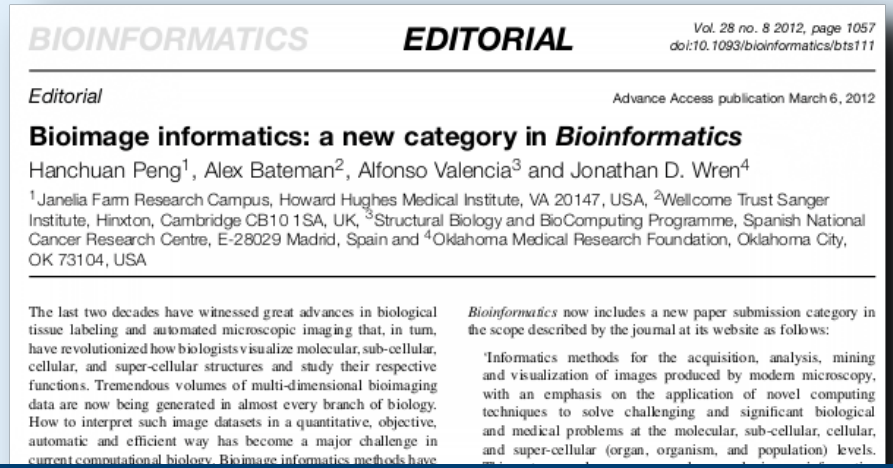
Leading Edge
Essay

Cell

Computer Vision in Cell Biology

Gaudenz Danuser[1,*]
[1]Harvard Medical School, 240 Longwood Avenue, Boston, MA 02140, USA
*Correspondence: gaudenz_danuser@hms.harvard.edu
DOI 10.1016/j.cell.2011.11.001

Computer vision refers to the theory and implementation of artificial systems that extract information from images to understand their content. Although computers are widely used by cell biologists for visualization and measurement, interpretation of image content, i.e., the selection of events worth observing and the definition of what they mean in terms of cellular mechanisms, is mostly left to human intuition. This Essay attempts to outline roles computer vision may play and should play in image-based studies of cellular life.



BIOINFORMATICS EDITORIAL Vol. 28 no. 8 2012, page 1057
doi:10.1093/bioinformatics/bts111

Editorial Advance Access publication March 6, 2012

Bioimage informatics: a new category in *Bioinformatics*

Hanchuan Peng[1], Alex Bateman[2], Alfonso Valencia[3] and Jonathan D. Wren[4]

[1]Janelia Farm Research Campus, Howard Hughes Medical Institute, VA 20147, USA, [2]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK, [3]Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre, E-28029 Madrid, Spain and [4]Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA

The last two decades have witnessed great advances in biological tissue labeling and automated microscopic imaging that, in turn, have revolutionized how biologists visualize molecular, sub-cellular, cellular, and super-cellular structures and study their respective functions. Tremendous volumes of multi-dimensional bioimaging data are now being generated in almost every branch of biology. How to interpret such image datasets in a quantitative, objective, automatic and efficient way has become a major challenge in current computational biology. Bioimage informatics methods have

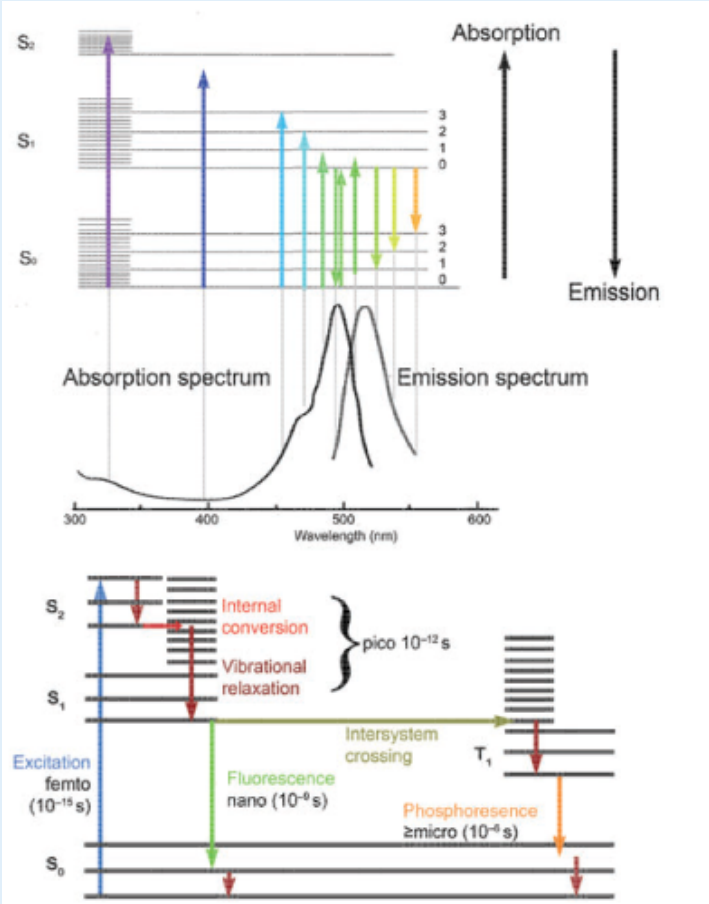
Bioinformatics now includes a new paper submission category in the scope described by the journal at its website as follows:

'Informatics methods for the acquisition, analysis, mining and visualization of images produced by modern microscopy, with an emphasis on the application of novel computing techniques to solve challenging and significant biological and medical problems at the molecular, sub-cellular, cellular, and super-cellular (organ, organism, and population) levels.

Fluorescence microscopy

Principle

- Fluorescent dyes are added to the sample.
- The dyes are excited by a source of illumination of a given wavelength
- The dyes are emitting at a longer wavelength and their response is registered by a CCD camera
- After several emission cycles, the dyes bleach



Lichtman J. and Conchello J.-A., Fluorescence Microscopy, *Nature Methods*, vol. 2, 2006

Images at single molecule resolution

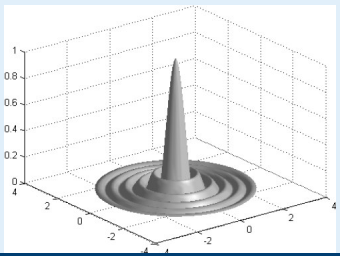
Point spread function

The image of a point source is the point spread function. Fluorophores can be regarded as point sources.

Theoretical models of PSF

- Richards-Wolf model
- Gibson-Lanni model

Approximations of PSF



"All models are wrong, but some are useful." (G. Box)

- Airy function

$$\mathrm{PSF}(r) = \left(2\frac{J_1(\pi q_c r)}{\pi q_c r}\right)^2,\ q_c = \frac{2NA}{\lambda}$$

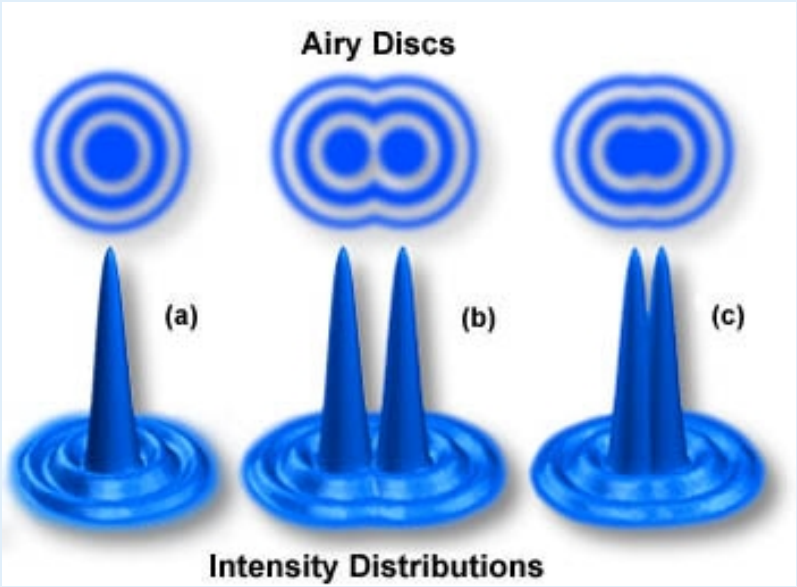
- Gaussian: $G(r) = e^{-\left(\frac{r^2}{2a^2}\right)}$
- modified Lorentzian: $L(r) = \frac{1}{1+\left(\frac{r^2}{a^2}\right)^b}$

Resolution limits

Rayleigh criterion

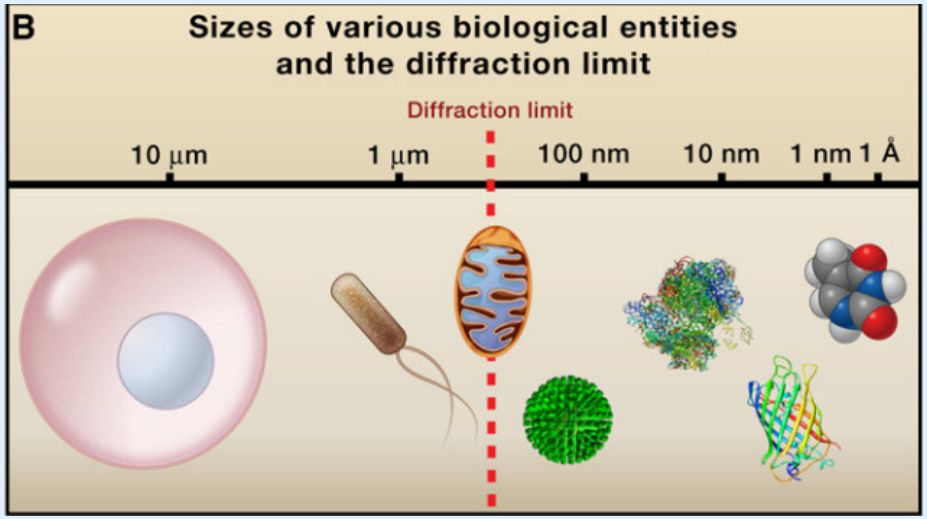
Two point sources are regarded as just resolved when the principal diffraction maximum of one image coincides with the first minimum of the other.

$$R = \frac{0.61\lambda}{NA}$$



Light Sheet Microscopy

Resolution scale

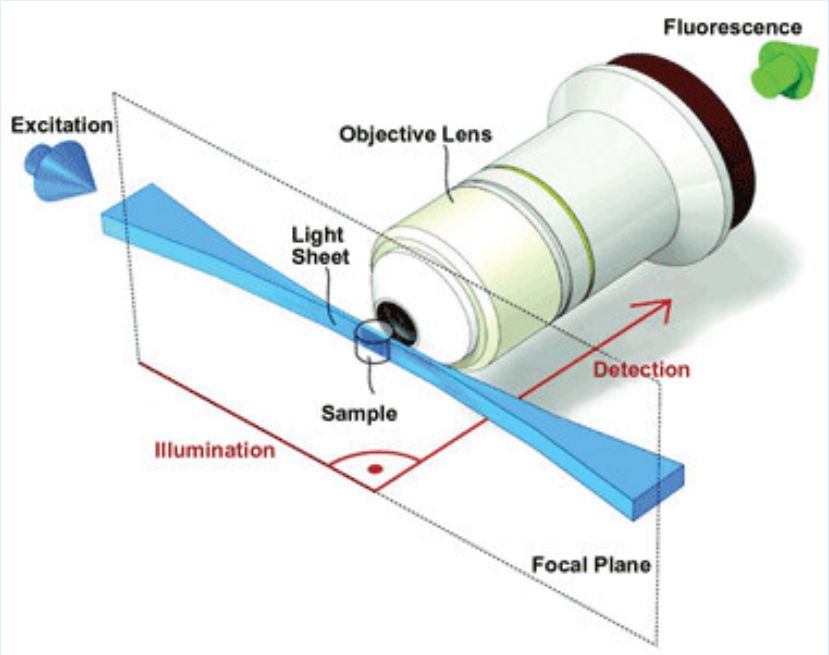


Left to right: A mammalian cell, a bacterial cell, a mitochondrion, an influenza virus, a ribosome, the green fluorescent protein, and a small molecule (thymine)

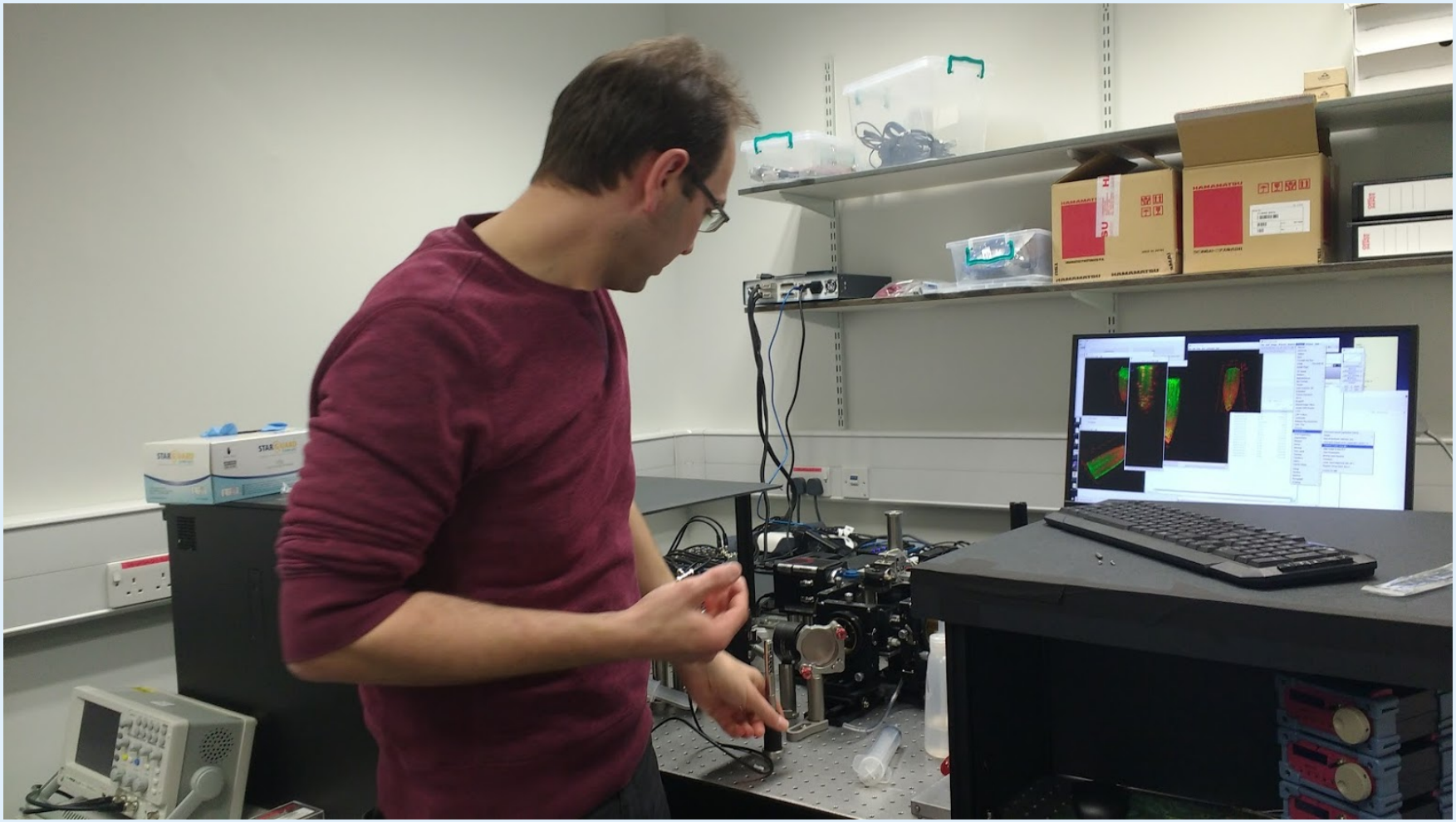
B. Huang, H. Babcock, and X. Zhuang - Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells, *Cell*, 143, 2010

Light sheet microscopy

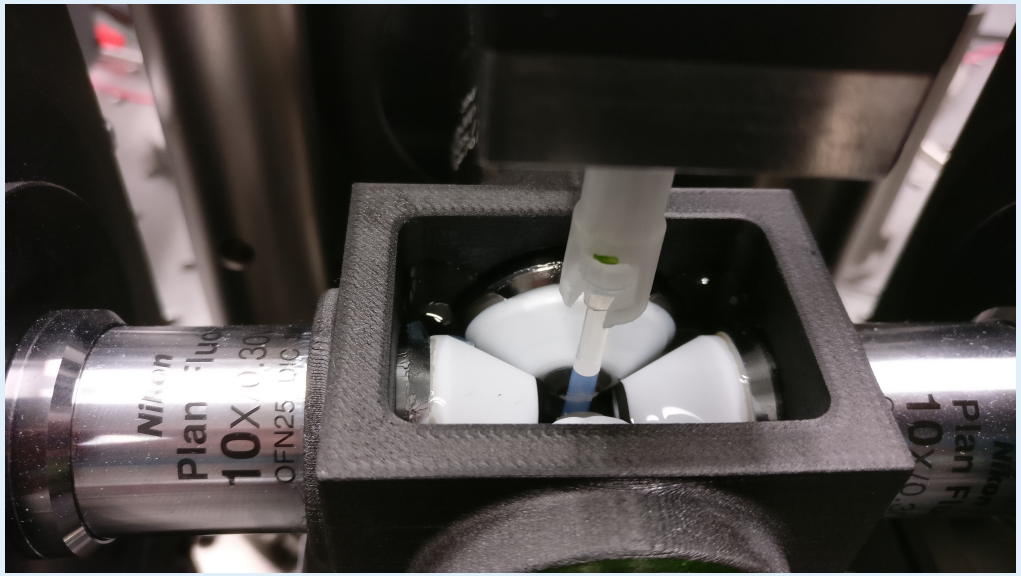
- *Nature methods* technique of 2014
- Fast imaging of large volumes, e.g. entire embryos imaged for several days
- Low photo-toxicity
- One experiment - terabytes of data (3D - 5D image stacks)



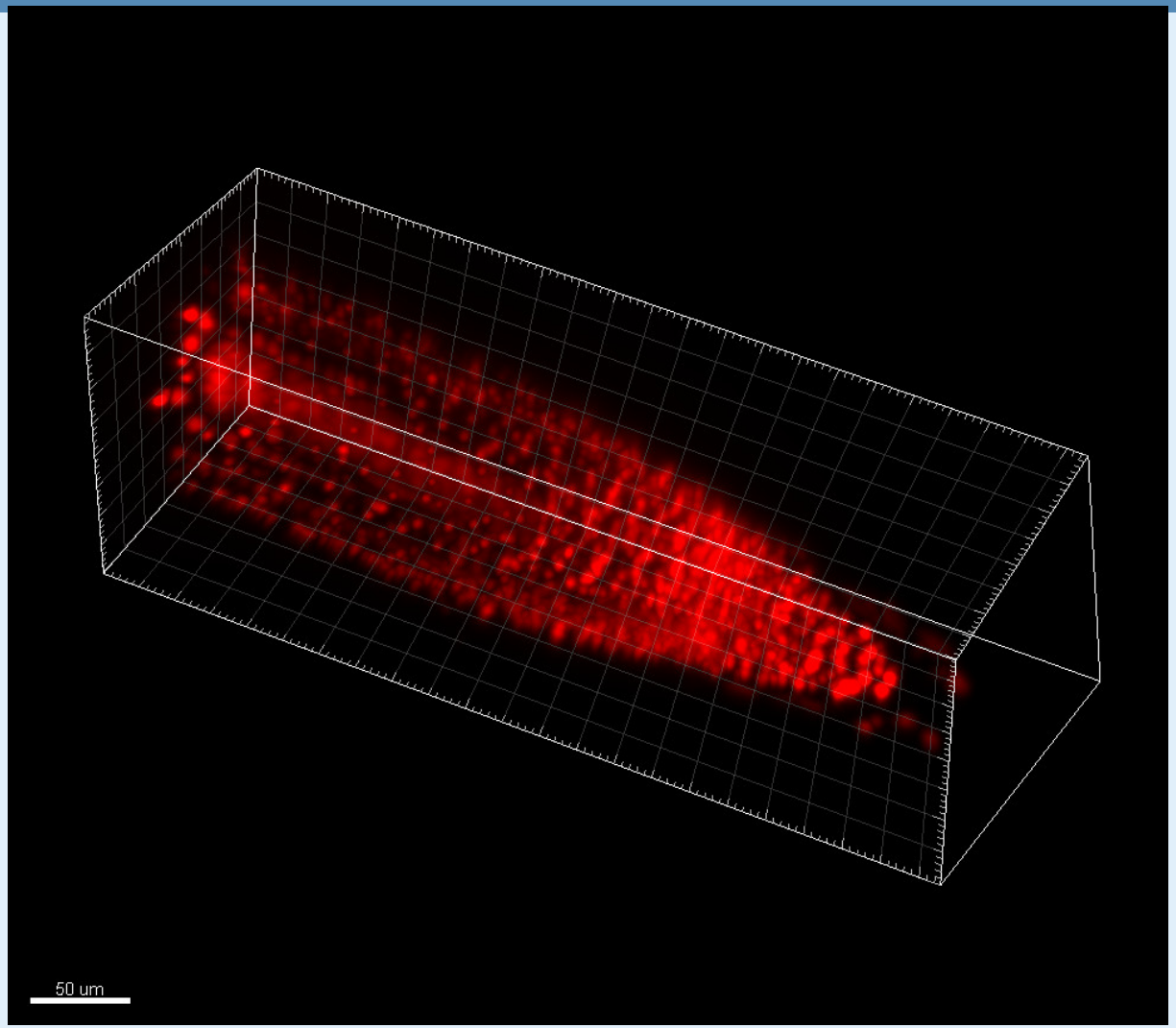
Martin and the SLCU system...



Sample preparation and mounting



First sample: *Arabidopsis thaliana* root

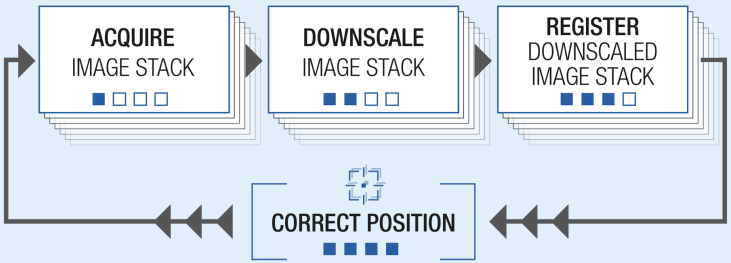


Katie Abley
Locke group

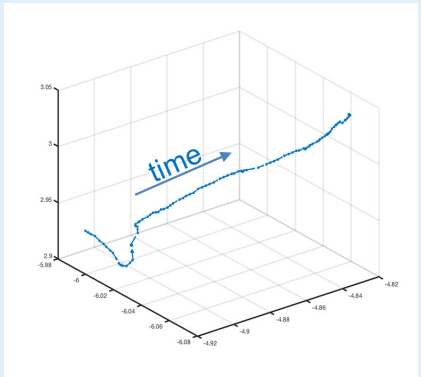
Imaris(Bitplane) visualisation

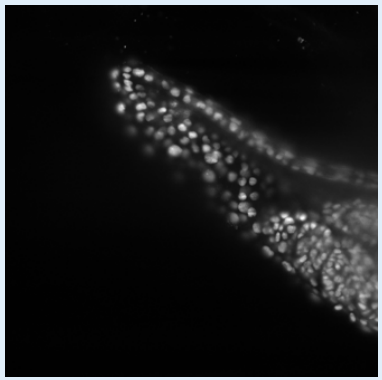
"Smart" imaging

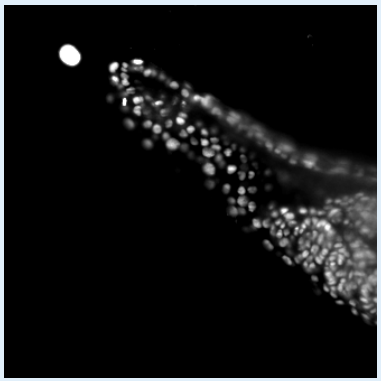
- The region of interest might exit the field of view (drift, growth of the sample)
- Correction offset from downsampled images (normalized correlation)



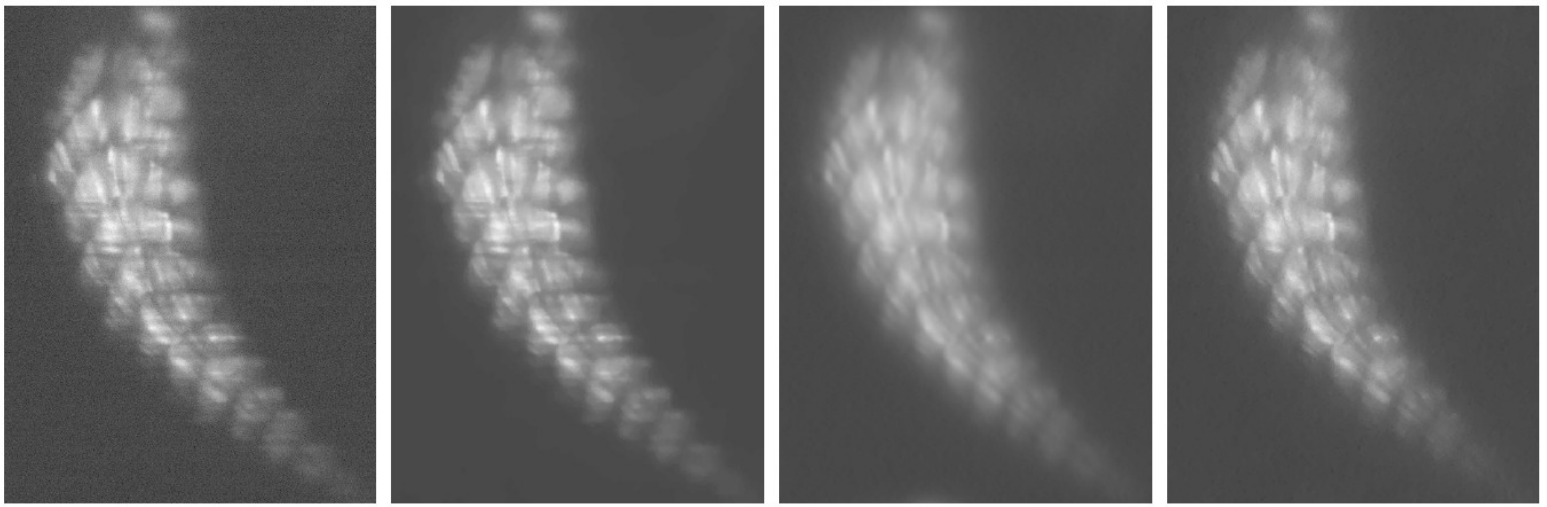
- Sample is brought in the field of view (by controlling the stage position)
- Robustness (artifacts might hamper the quality of the registration)





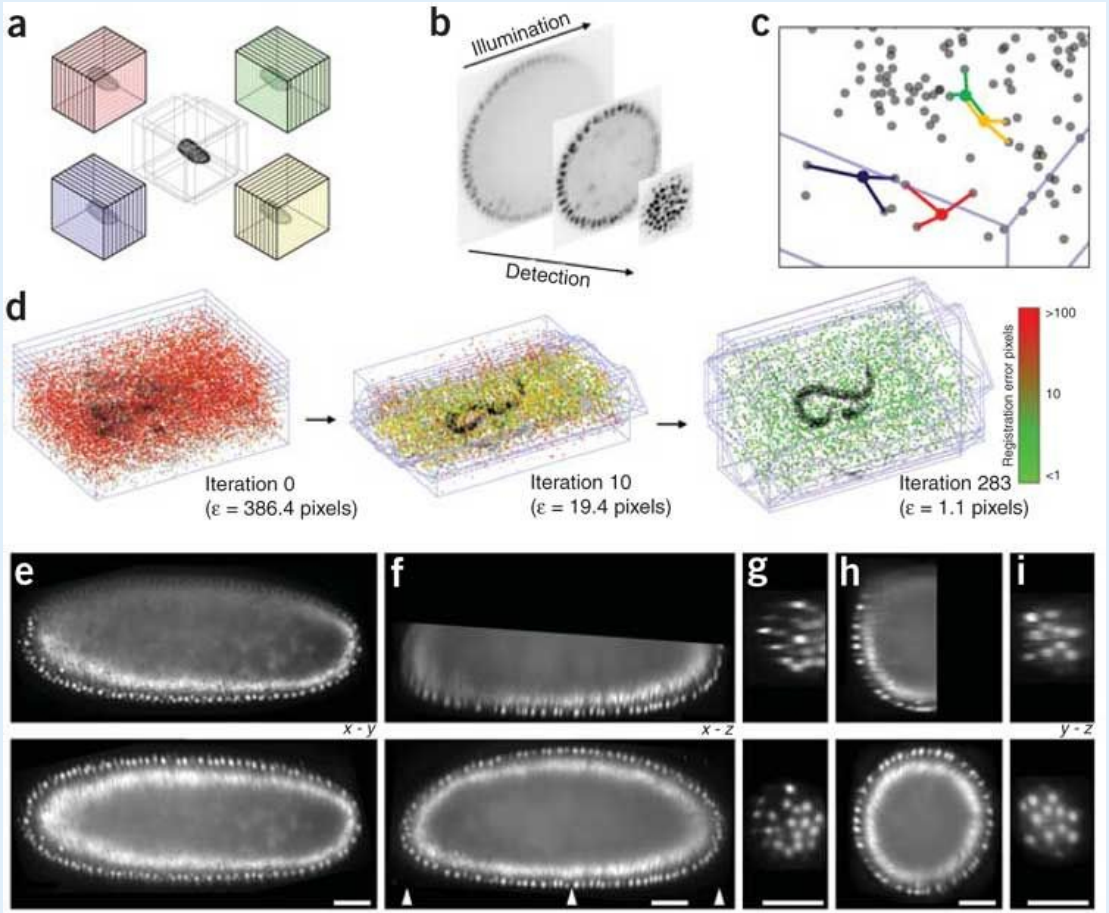


Removing scattering artefacts



Variational algorithms to remove stationary noise. Application to microscopy imaging. J. Fehrenbach, P. Weiss and C. Lorenzo, IEEE Image Processing Vol. 21, Issue 10, pages 4420 - 4430 (2012)

Multi-view fusion / Registration



Software for bead-based registration of selective plane illumination microscopy data Preibisch S., Saalfeld S., Schindelin J., Tomancak P, Nature Methods volume 7 (2010)

Tracking of cells in time

Challenges:

- Large data sets
- Quality of data varies in the sample
- Densely packed nuclei
- Noisy, blurred data

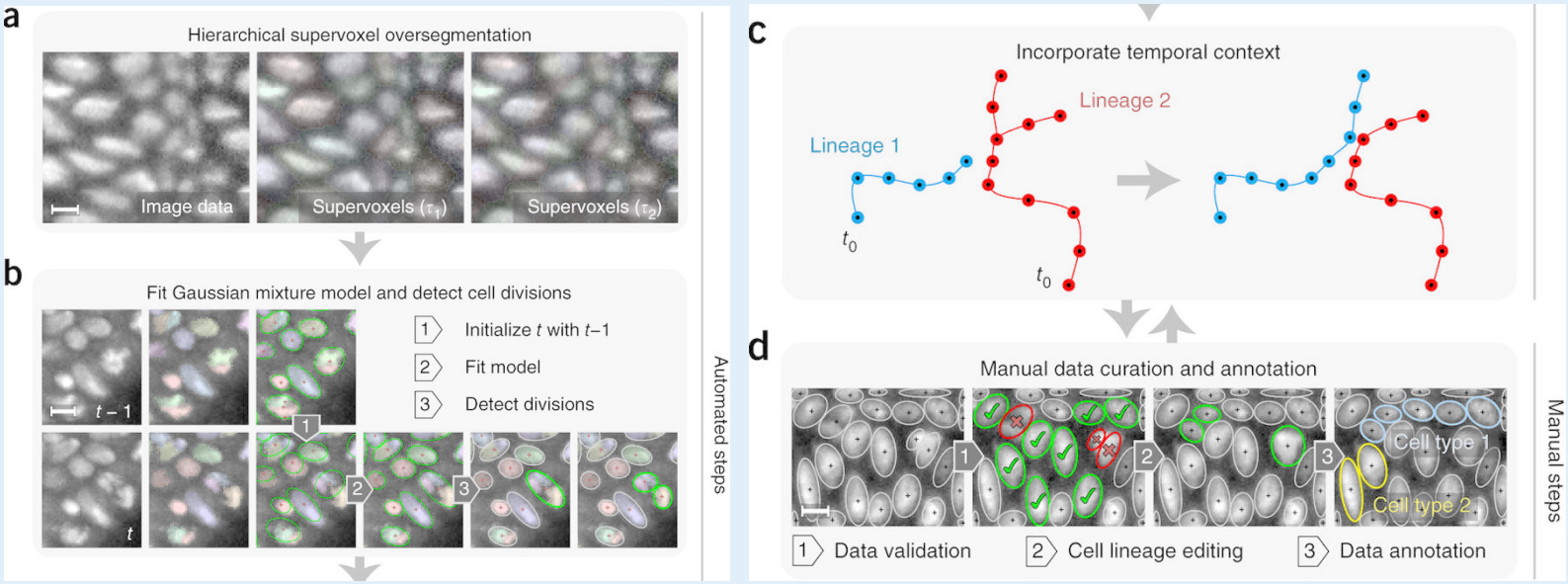
Software:

- Imaris (Bitplane)
- TGMM (Keller lab, Janelia)
- Bioemergences (CNRS, Gif-sur-Yvette)

Tracking: TGMM software

- generality by reconstructing cell lineages in four-dimensional, terabytesized image data sets
- up to 20,000 cells per time point (26,000 cells / min) on a single computer workstation and
- ease of use (two parameters)
- visualization and editing tools for efficient data curation.

Tracking: TGMM software



Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data
Fernando Amat, William Lemon, Daniel P Mossing, Katie McDole, Yinan Wan, Kristin Branson, Eugene W Myers, Philipp J Keller

Nature Methods (2014) doi:10.1038/nmeth.3036

Optimized HPC pipeline, hyper-parameter search, cell division detection

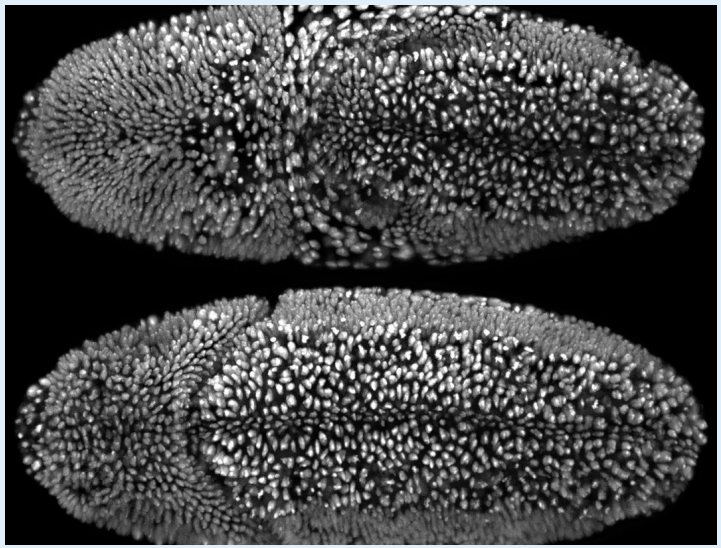
Data management challenges

- Huge data sets to store, analyze, share: minimize size, number of copies
  - What is "raw data"?
- Complex additional information: management of metadata, results, annotations
- Data accessibility

Reducing data size: Map projections

Embryos of major model organisms (zebrafish, frog, fruit fly and so on) exhibit a spherical or ellipsoidal shape.

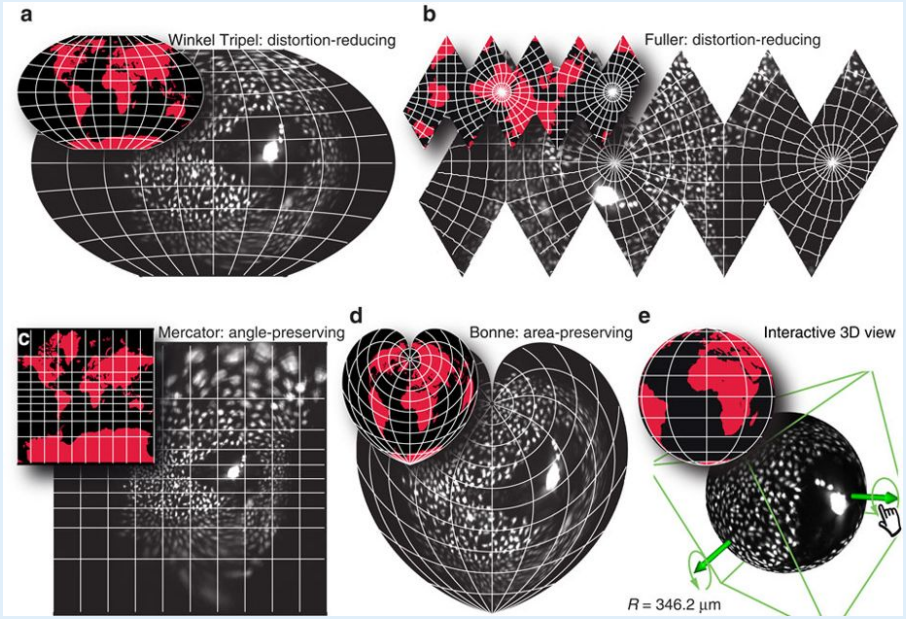
- Cuboidal images - inefficient representation
- Image transformation on the fly : $3D \rightarrow 2D$
- Analysis and visualization become straightforward.
- Huisken group - Original example: the endoderm (monolayer on the surface of the spherical yolk) during early stages of zebrafish development



Amat, F. et al - Efficient processing and analysis of large-scale light-sheet microscopy data, *Nature Protocols*, 10 (1679), 2015

Reducing data size: Map projections

- (a,b) Winkel Tripel and Fuller projections: little distortion - speed of cell migration.
- (c) Mercator projections: angle-preserving - cell flows and direction of cell movement.
- (d) The Bonne projection: area preserving - cell densities.
- (e) Cell dynamics visualized interactively in 3D viewer (Fiji)

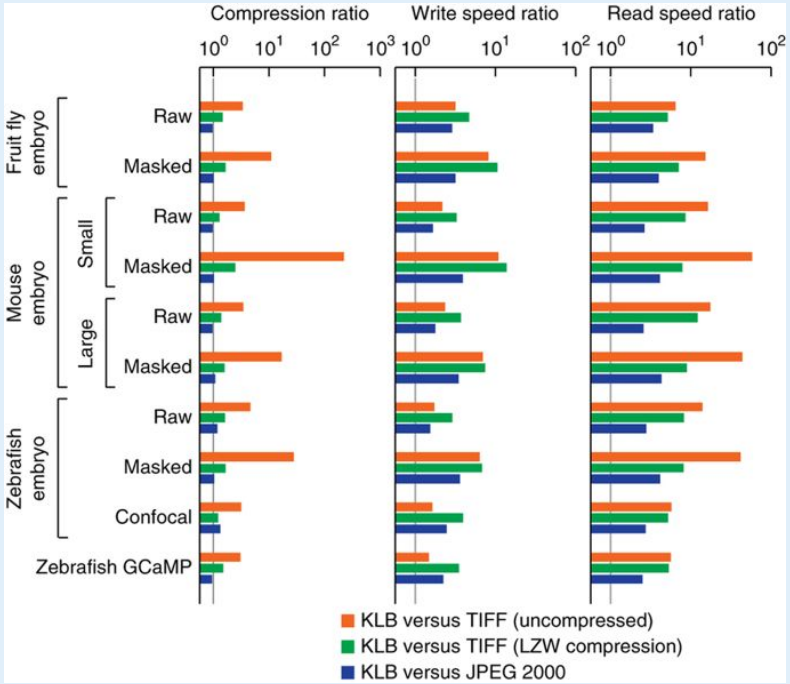


Schmid, B., et al. - High-speed panoramic light-sheet microscopy reveals global endodermal cell dynamics, *Nature Communications*, 4 (2207), 2013

Reducing data size: KLB format

Keller Lab Block Amat, F. et al - Efficient processing and analysis of large-scale light-sheet microscopy data, *Nature Protocols*, 10 (1679), 2015

- lossless image-compression format
- high compression ratios, fast read/write speeds, flexible block architecture
- partitions in 5D blocks and compresses blocks in parallel using BZip2
- both reading and writing operations are parallelized, they scale linearly with the number of cores in the CPU
- API for interfacing the open-source C++ code with various platforms



Managing data

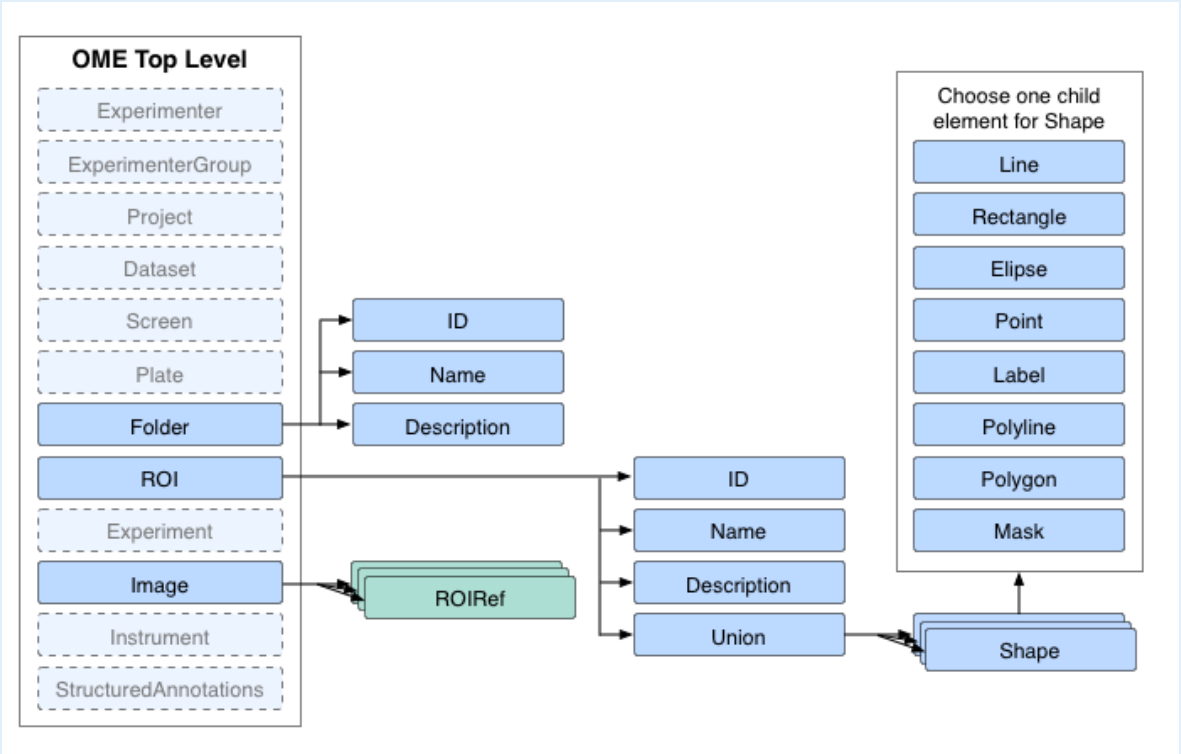
Data = images + all related information (metadata)

Open Microscopy Environment `openmicroscopy.org`

Jason Swedlow, University of Dundee / Eurobioimaging /EMBL-EBI

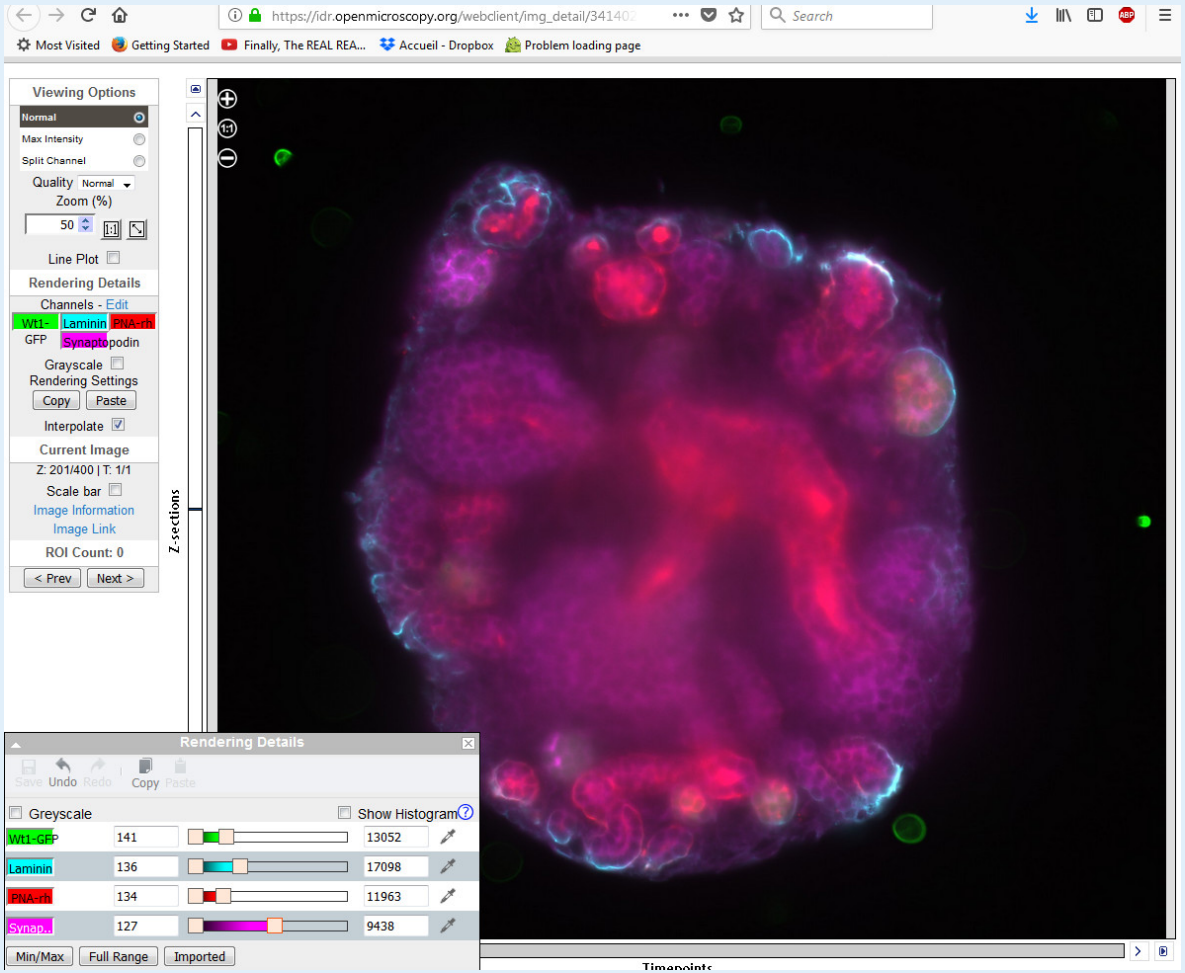
- OME data model, **OME-TIFF** (TIFF or BigTIFF file with OME-XML metadata block describing the dataset embedded in header) *.ome.tif, .ome.btf,...*
- **Bio-formats** - can read/write and convert more than 145 file format types to OME-TIFF
- **OMERO** -modern client-server software platform for visualizing, managing, and annotating scientific image data

Managing data



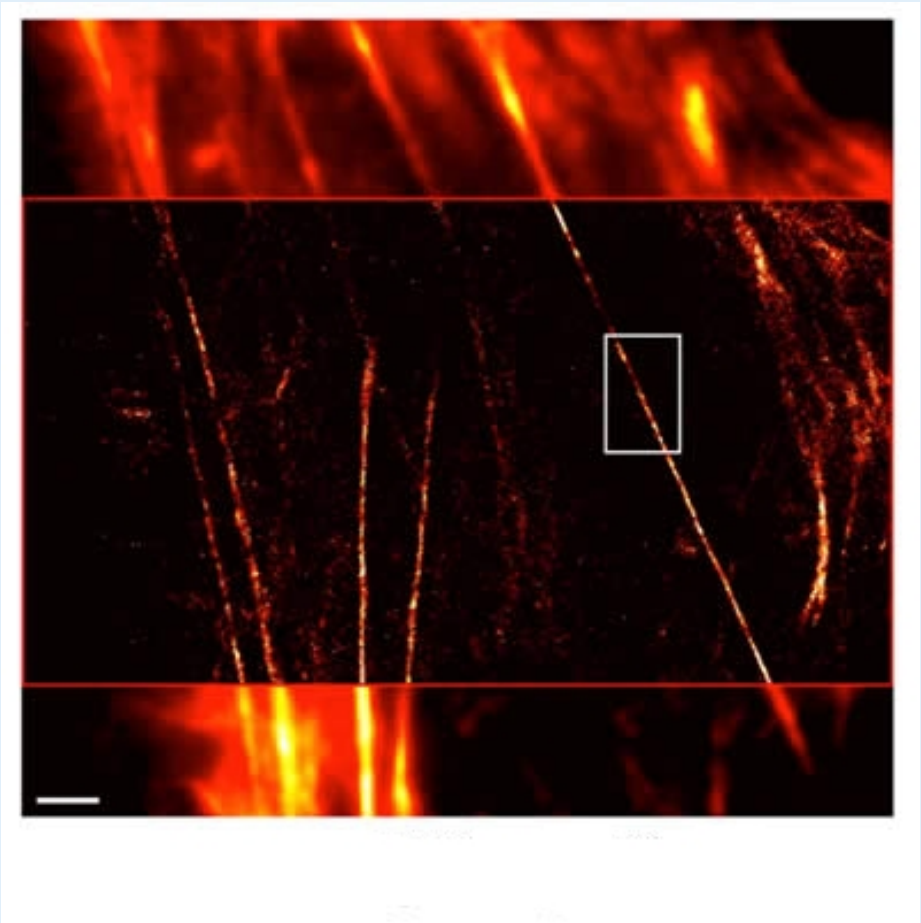
Sharing data: FAIR principles

Image Data Repository (IDR) - `idr.openmicroscopy.org`



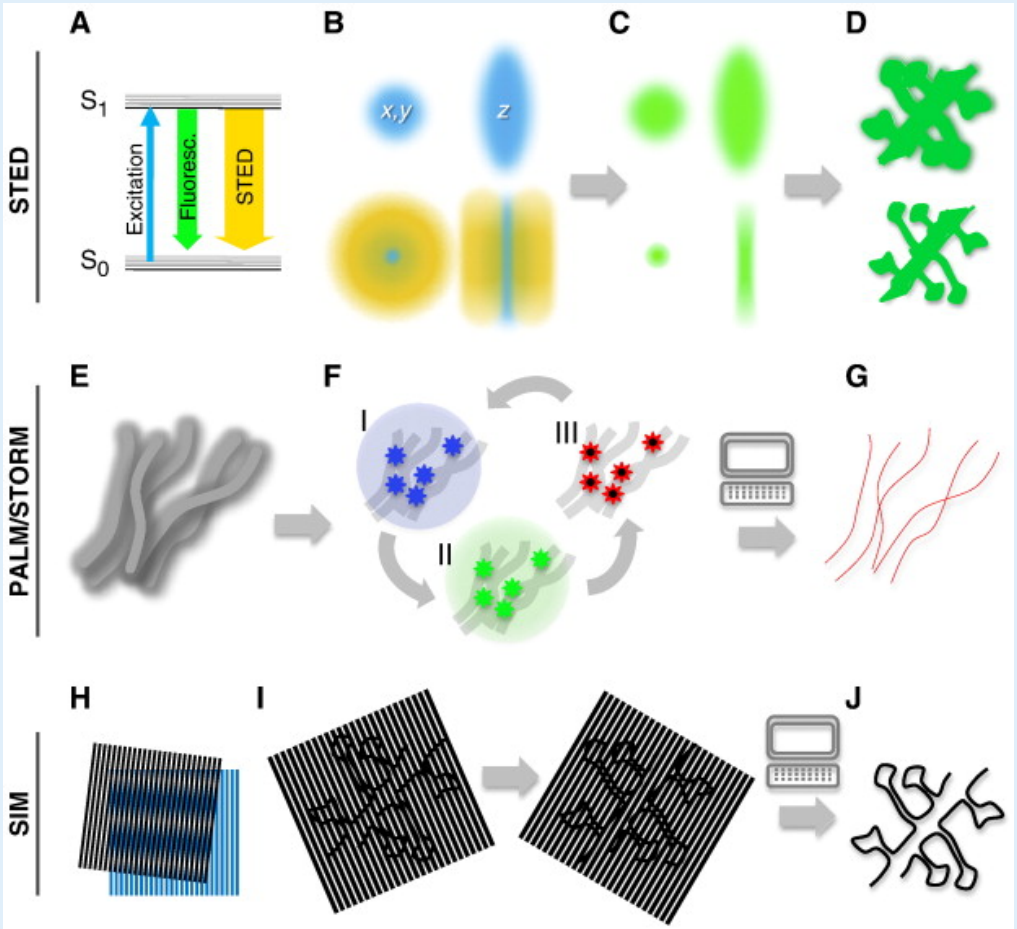
Super-resolution light microscopy

Goal: increased resolution



P. N. Hedde, et. al. Nature Methods 6, 689-690 (2009).

Summary of superresolution techniques



J. Tonnesen, U. Valentin Naegerl, Superresolution imaging for neuroscience, Experimental Neurology, 2010

Nobel Prize in Chemistry 2014

Photo: A. Mahmoud

Eric Betzig

Prize share: 1/3

Photo: A. Mahmoud

Stefan W. Hell

Prize share: 1/3

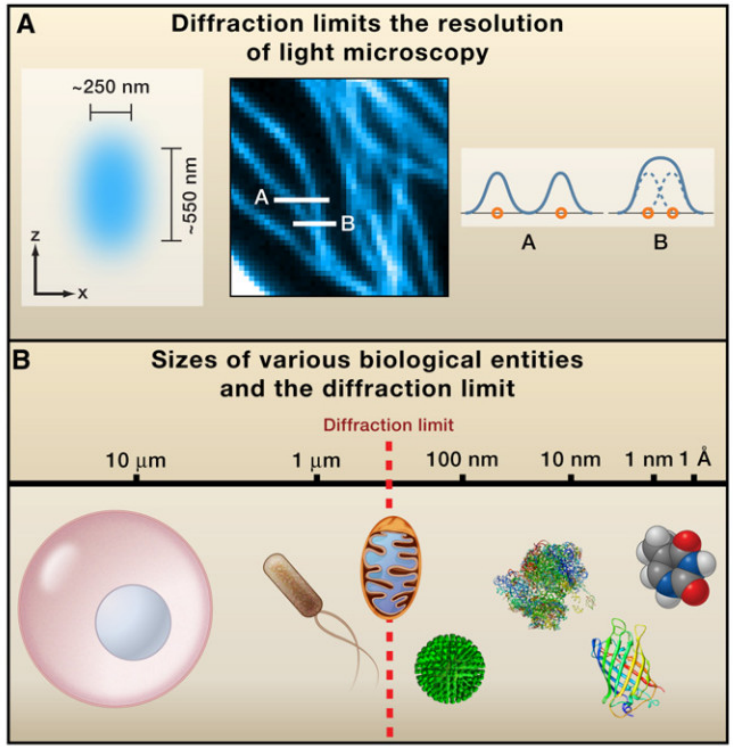
Photo: A. Mahmoud

William E. Moerner

Prize share: 1/3

The Nobel Prize in Chemistry 2014 was awarded jointly to Eric Betzig, Stefan W. Hell and William E. Moerner *"for the development of super-resolved fluorescence microscopy"*.

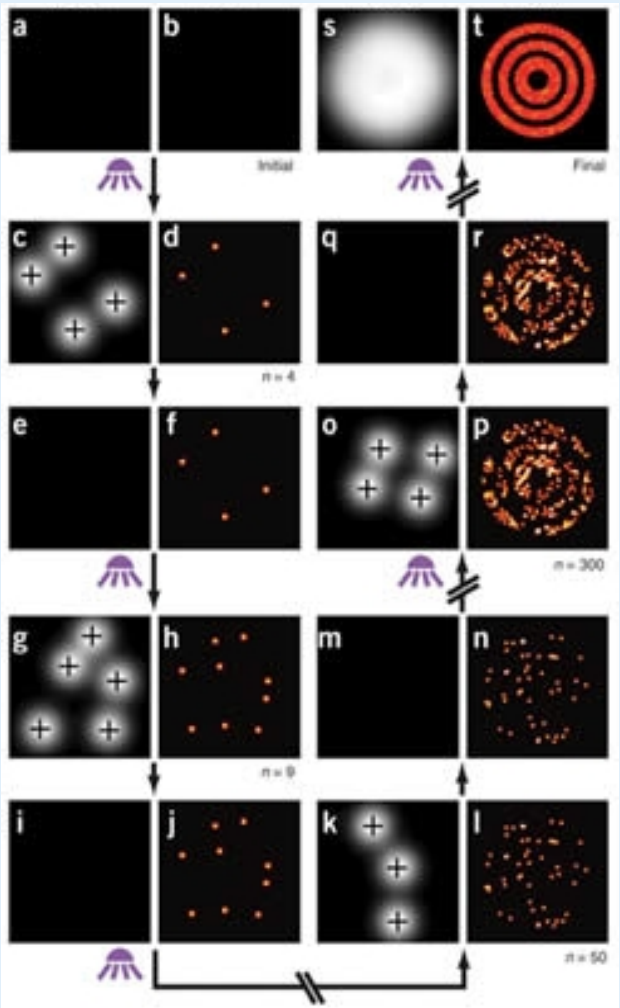
Resolution scale



Left to right: A mammalian cell, a bacterial cell, a mitochondrion, an influenza virus, a ribosome, the green fluorescent protein, and a small molecule (thymine)

B. Huang, H. Babcock, and X. Zhuang - Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells, *Cell*, 143, 2010

What is "nanoscopy"?



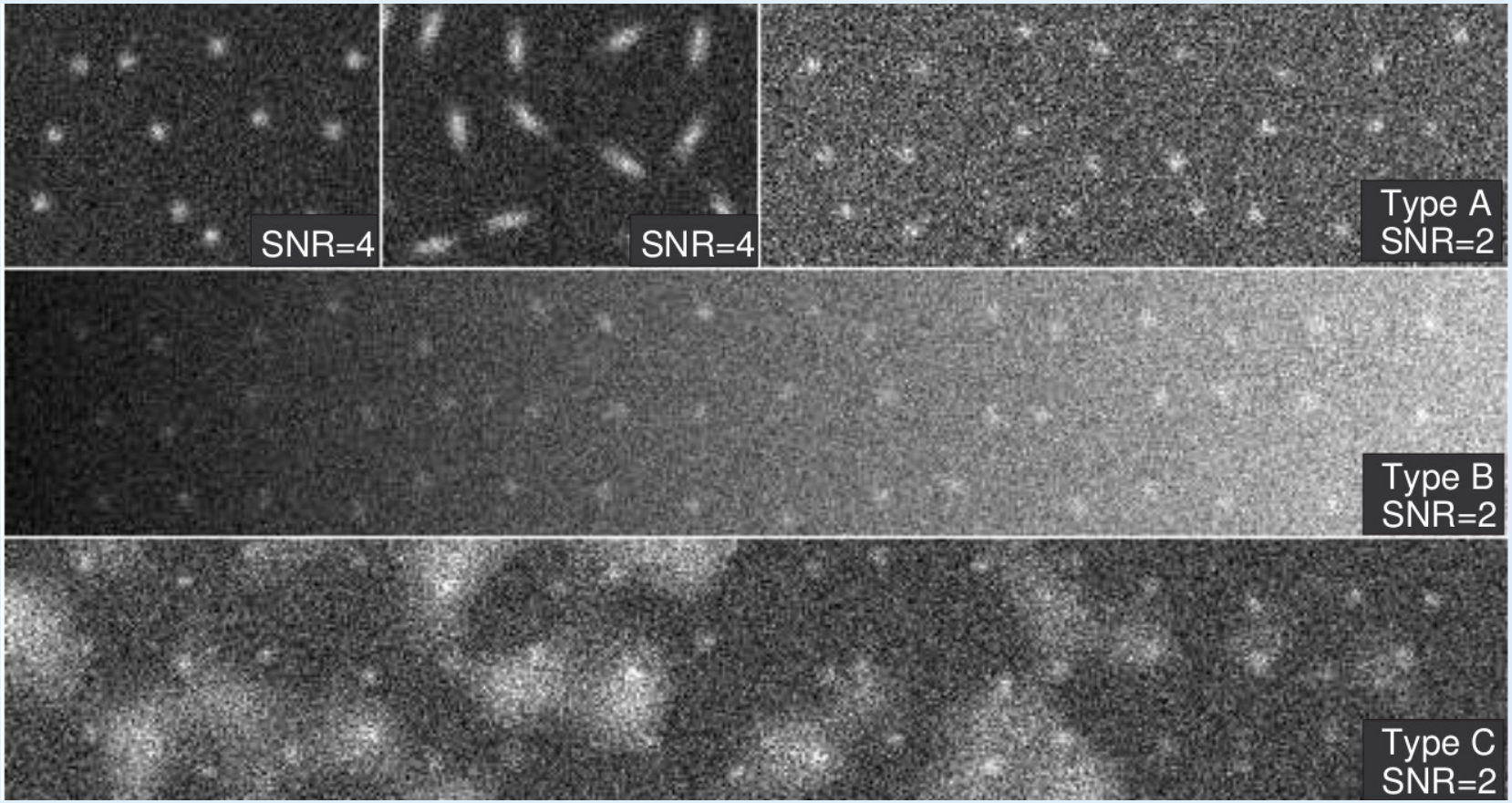
Characteristics

- Fluorophores stochastically alternating between active and dark state
- Algorithms localizing individual fluorophores in images

T. J. Gould, V. V. Verkhusha, S.T. Hess, Nature Protocols

Detection and fitting of single molecules

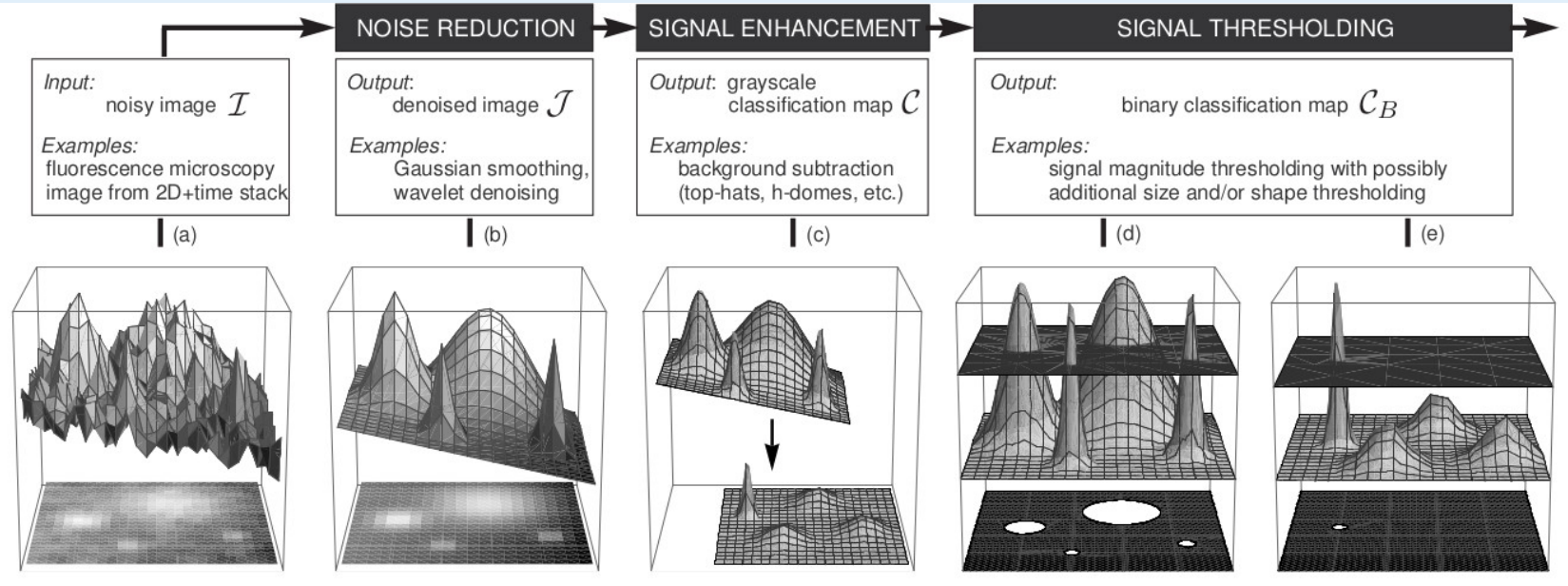
Challenges



Ihor Smal et al., *Quantitative Comparison of Spot Detection Methods in Fluorescence Microscopy*, IEEE TMI, 2012

Detection and fitting of single molecules

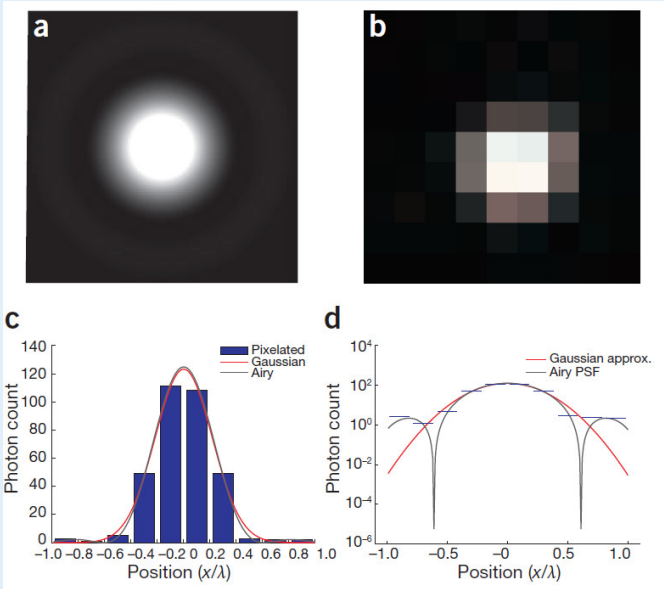
Framework for detection



Ihor Smal et al., *Quantitative Comparison of Spot Detection Methods in Fluorescence Microscopy*, IEEE TMI, 2012

Detection and fitting of single molecules

Fitting the PSF model at the detected position



Small A. and Stahlheber S.- Fluorophore localization algorithms for super-resolution microscopy, *Nature Methods*, vol.11, no.3, 2014

Model:

$$I(x, y) = I_0 h(x - x_0, y - y_0) + b$$

Fitting

1. Least squares minimization

$$S = \sum_{pixels} \frac{(data - model)^2}{var(data)}$$

2. Maximum likelihood estimation (based on image formation model)
3. Non-iterative approaches: CoG, radial symmetry based algorithm, triangulations etc.

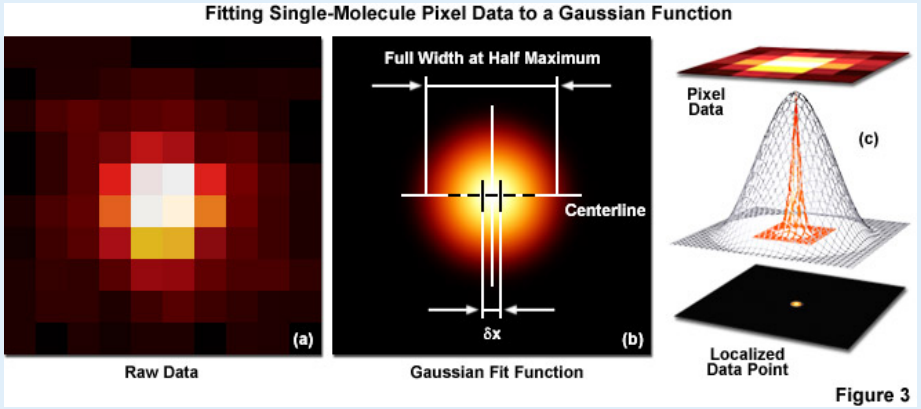
Fitting of PSF models

Factors that affect precision:

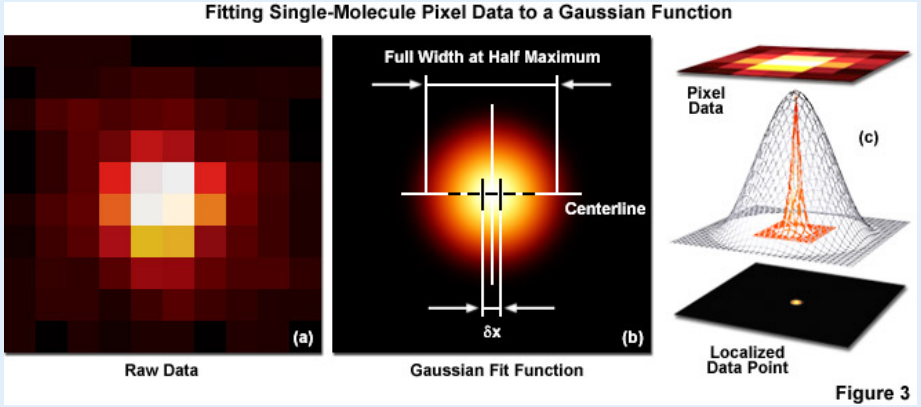
- Model imprecision
- Background (from out-of focus fluorophores e.g.)
- Noise
- Spot density

Noise

- Shot noise (Poisson)
- Read-out noise (Gaussian)
- EMCCD (heteroscedastic)



Fitting of PSF models



In the case of MLE the precision of the estimate is known:
Cramér-Rao lower bound (inverse of the Fisher information) $\mathrm{Var}(\hat{\theta}) \geq 1/I(\theta)$

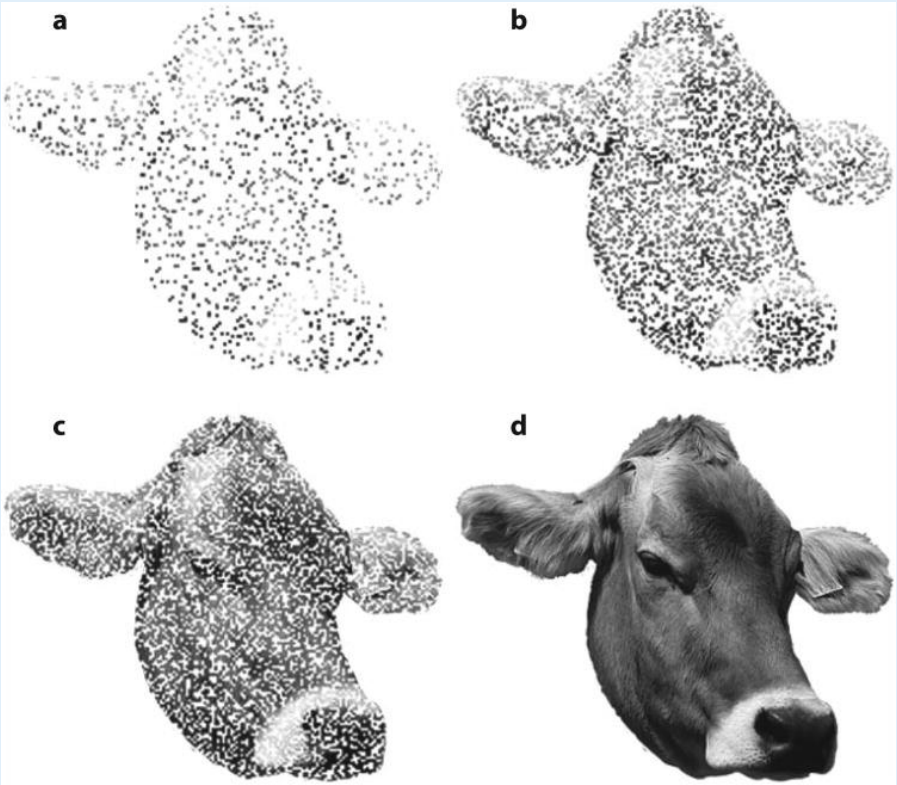
$$\mathrm{Var}(x) = \frac{\sigma_a^2}{N}\left(\frac{16}{9} + \frac{8\pi\sigma_a^2 b^2}{N}\right)$$

where $\sigma_a^2 = \sigma^2 + a^2/12$, a pixel witdh, b background intensity and N photon count. (Mortensen et al., Nature Methods, 7, 377 - 381, 2010).

Localisation precision: $\approx 2 - 25$ nm

Resolution

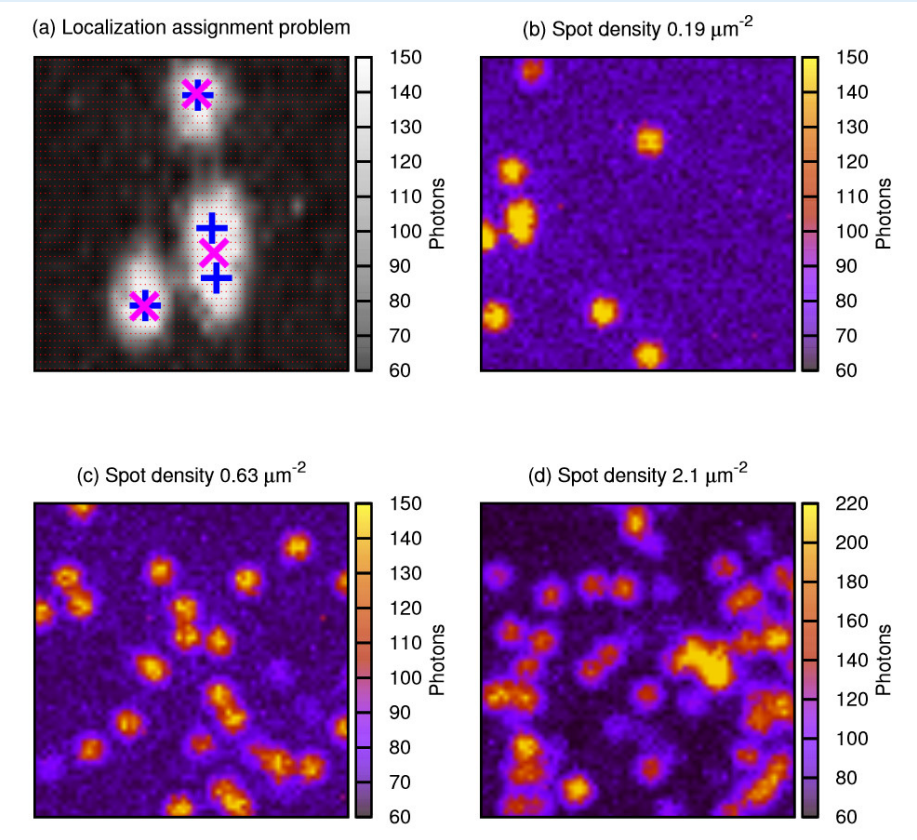
Fitting precision + Density of labeling



Patterson G., Davidson M., Manley S., Lippincott-Schwartz J. - Superresolution imaging using single-molecule localization. *Ann. Rev. Phys. Chem.* 61:345-67, 2010

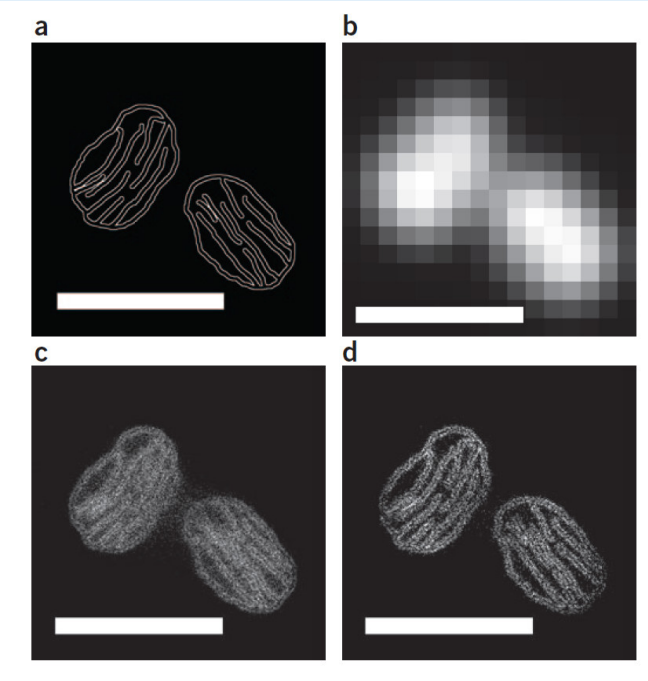
High density images: multiple overlapping spots

Necessary when density of spots increases



Wolter S., Endesfelder U., van de Linde S., Heilemann M. and Sauer M. - Measuring localization performance of super-resolution algorithms on very active samples, Optics Express, Vol. 19, No. 8, 2010

Multiple overlapping spots

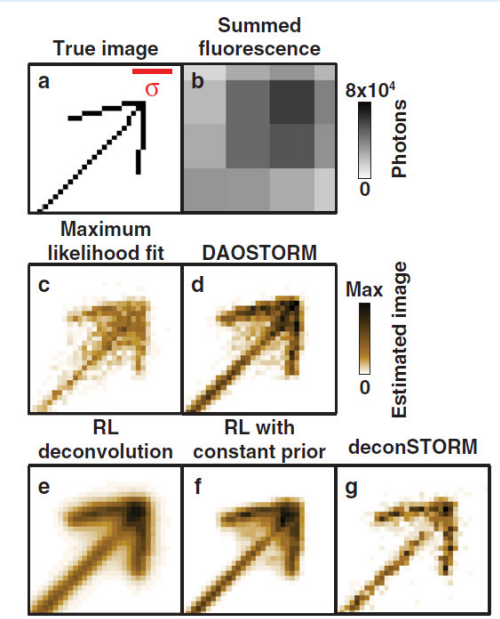


a) Simulated ground truth image
b) Sum of all fluorescence images
c) Result of single molecule fit algorithm
d) Result of single molecule fit algorithm where spots not fulfilling certain quality criteria (e.g. circularity) were eliminated

Small A. and Stahlheber S.- Fluorophore localization algorithms for super-resolution microscopy, *Nature Methods*, vol.11, no.3, 2014

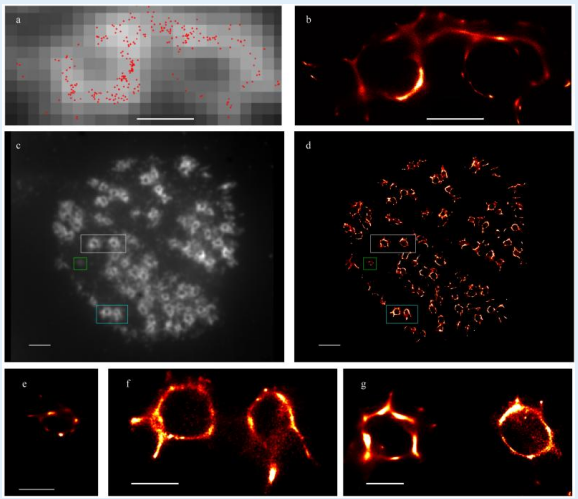
High density images

- Fitting multi-fluorophore models: DAOSTORM, PALMER etc.



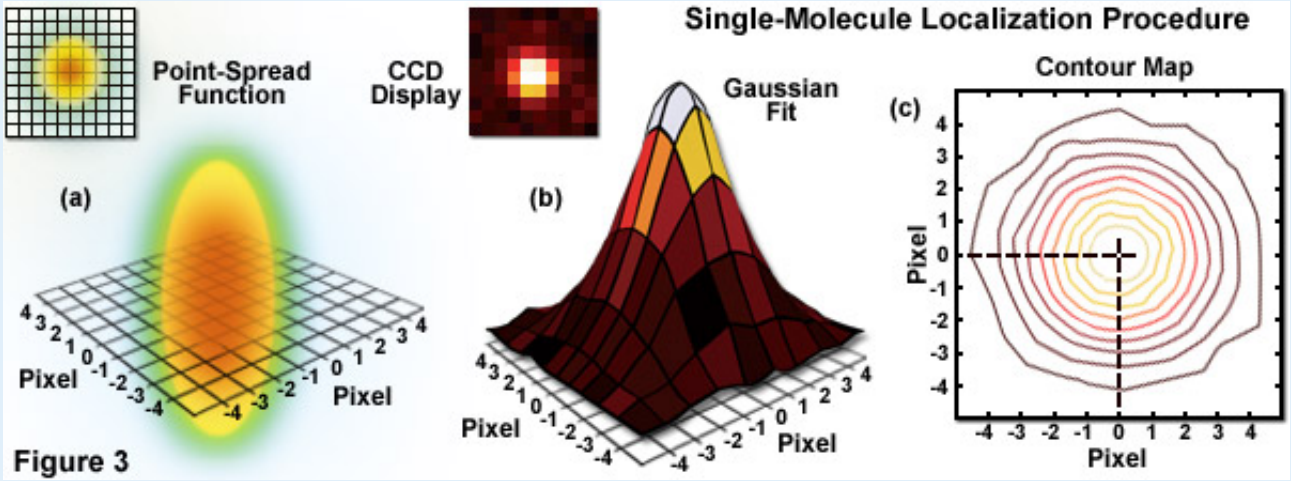
Mukamel E.A., Babcock H. and Zhuang X. - Statistical Deconvolution for Superresolution Fluorescence Microscopy, *Biophysical Journal*, Volume 102, 2012

- Fluorophore density estimation
  - Deconvolution
  - L_0/L_1 minimization
  - Bayesian estimation (3B)



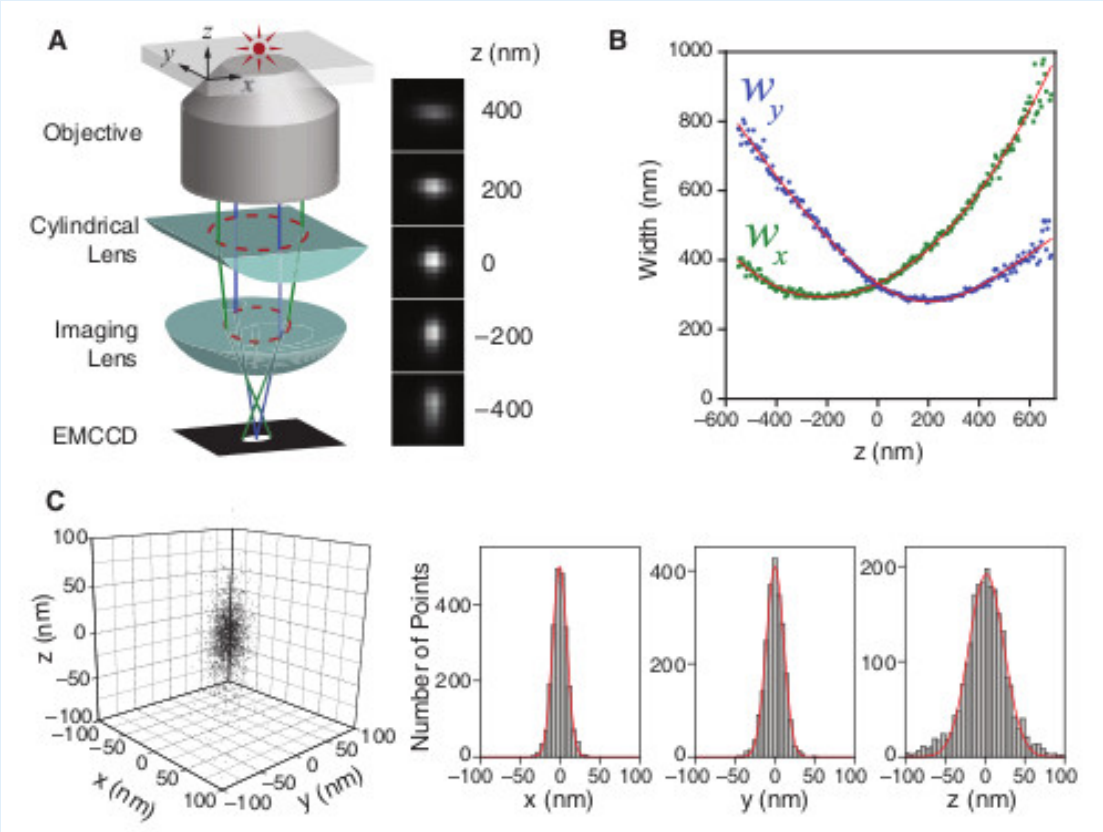
Cox S., Rosten E. et al. - Bayesian localisation microscopy reveals nanoscale podosome dynamics, *Nat Methods.* 2011, 9(2): 195 - 200.

3d superresolution



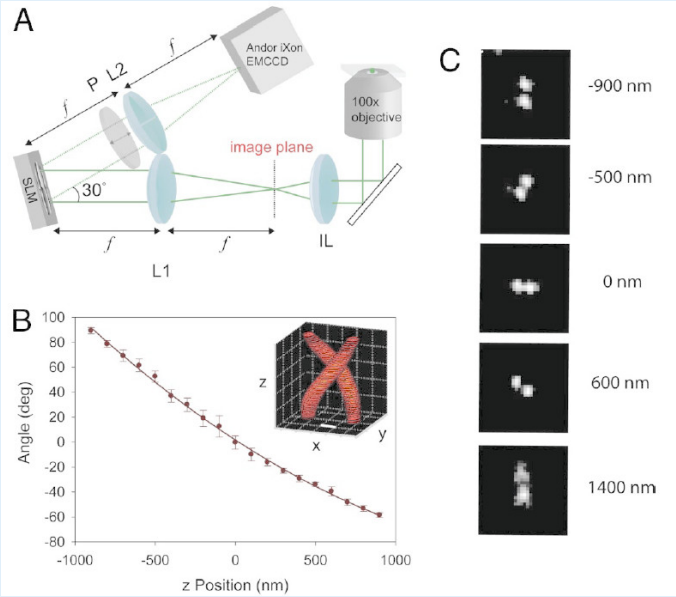
Zeiss-campus

3d superresolution via a cylindrical lens



B. Huang, W. Wang, M. Bates, X. Zhuang, "Three-dimensional Super-resolution Imaging by Stochastic Optical Reconstruction Microscopy", Science 319, 810-813 (2008)

Double helix PSF: Image formation



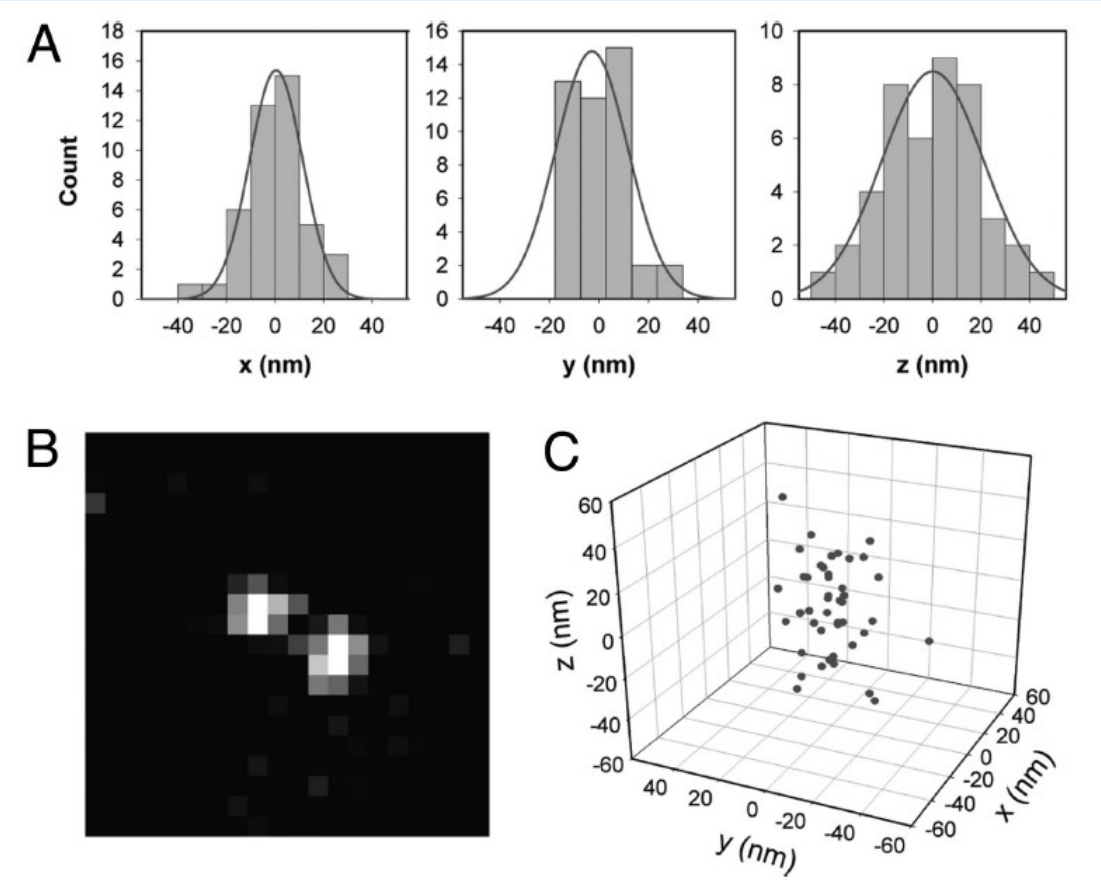
Pavani SRP, Thompson MA, ..., Moerner WA - Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function, PNAS, vol. 106, no. 9, 2009.

- spatially rotating point-spread functions (depth from diffracted rotation)
- inspired from depth from defocus techniques (circular aperture)

A SM near focus appears as two spots on a detector:

- Lateral (x, y) position - the midpoint between spots
- Axial(z) position - the angle of the line connecting the two spots and a fixed orientation (calibration measurement).

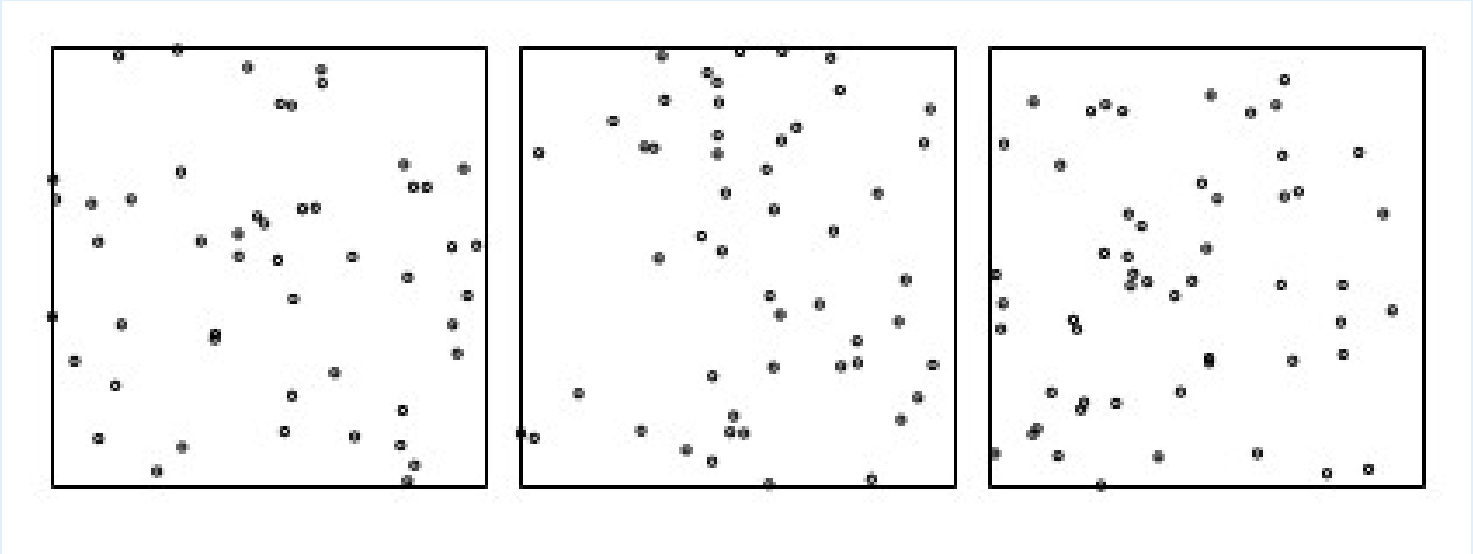
Double helix PSF: Precision



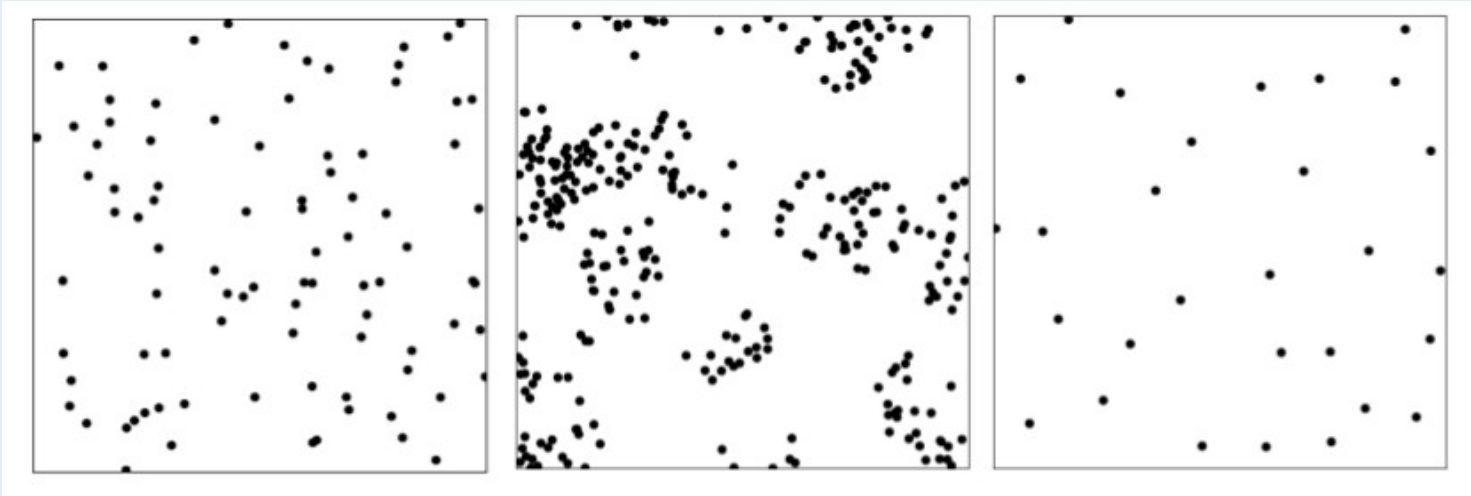
Pavani SRP, Thompson MA, Biteen JS, Lord SJ, Liu N., Twieg R.J, Piestun R. and Moerner W. E. - Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function, PNAS, vol. 106, no. 9, 2009.

Motivation

Spot the differences between these three pictures



Motivation



Summary statistics (second-order)

Ripley's K function:

$$K(r) = \frac{\mathbb{E}_o(N(b(o,r)\backslash\{o\})}{\lambda}$$

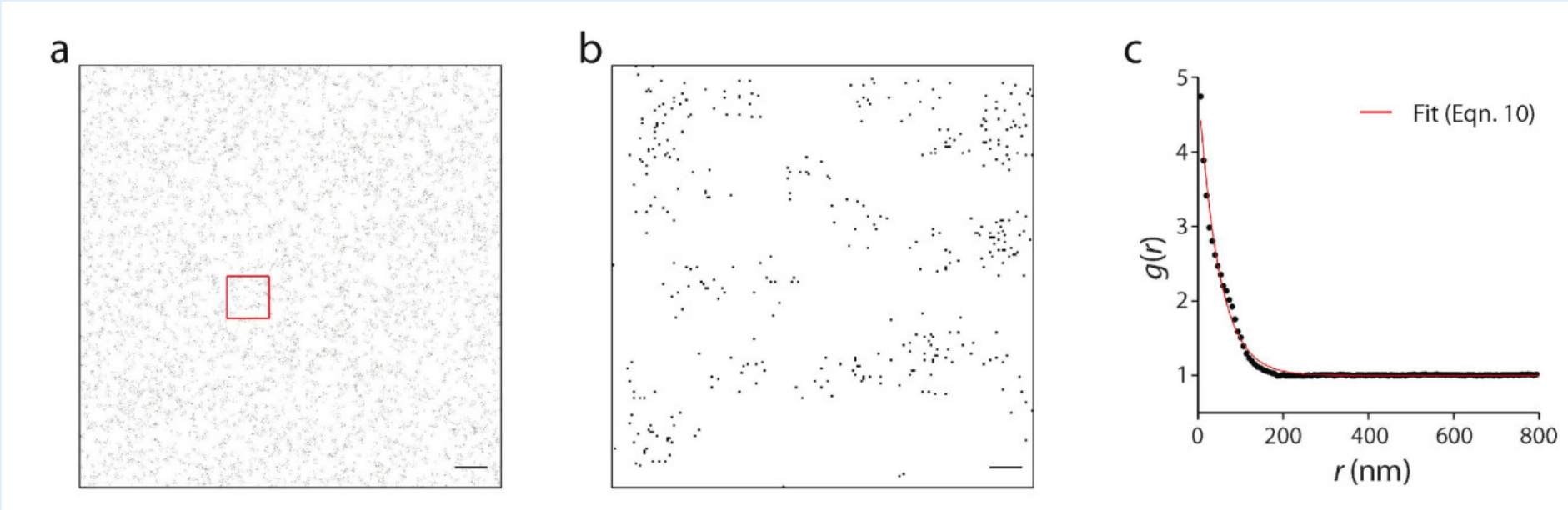
The mean number of points in a disc of radius r at the typical point (which is not counted).

Pair correlation function:

$$g(r) = \frac{1}{2\pi r}\frac{\mathrm{d}\,K(r)}{\mathrm{d}\,r} = \frac{\rho^{(2)}(r)}{\lambda^2}.$$

- For Poisson process: $K(r) = \pi r^2, \quad g(r) = 1.$
- For regular processes: $K(r) < \pi r^2, \quad g(r) < 1.$
- Cluster processes: $K(r) > \pi r^2, \quad g(r) > 1.$

Summary statistics (second-order)



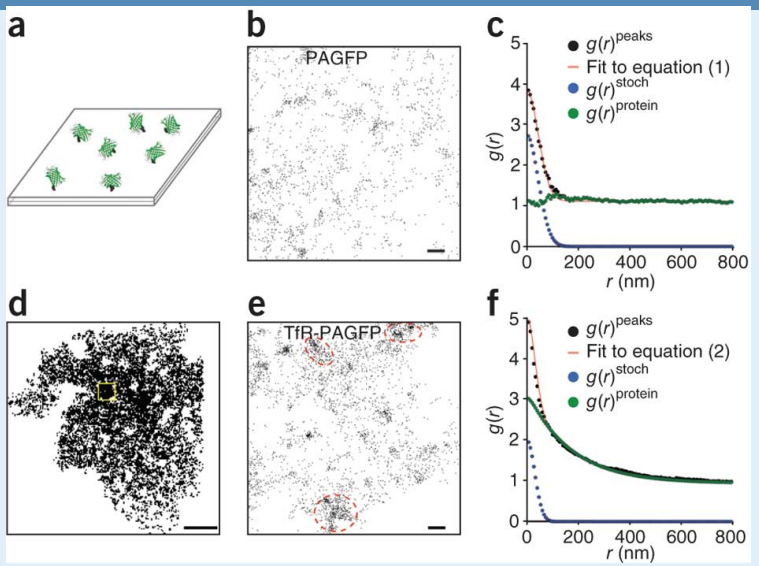
Protein heterogeneity in the plasma membrane using PALM and pair correlation analysis

- Combining pair-correlation analysis with PALM to analyze patterns of protein organization across the plasma membrane in COS-7 cells
- Models

$$g(r) = g(r)^{stoch} + 1 \quad (1)$$

$$g(r) = g(r)^{stoch} + + (A\exp(-r/\xi) + 1) * g(r)^{psf} \quad (2)$$

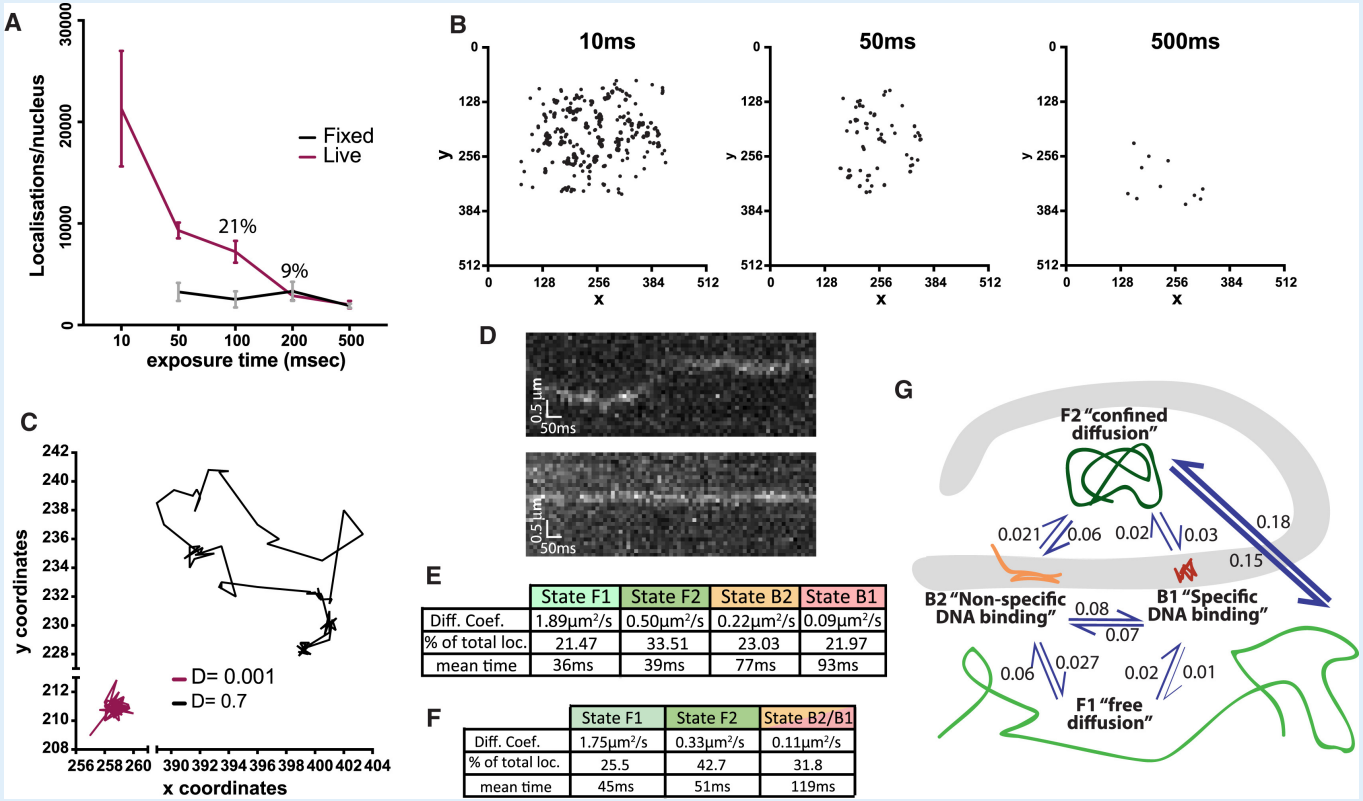
- Proteins:
  outer leaflet GPI-anchored protein, a transmembrane prot. (Lat), an inner leaflet lipid-anchored prot. (Lyn) and the transmembrane prot. vesicular stomatitis viral glycoprotein (VSVG).



(a,d) PAGFP molecules on coverslip and across the plasma membrane of a COS-7 cell (b,e) Localizations in a section of (a,d). (c,f) Fit of g(r) (to eq1 and eq2). Noticeable grouping of clusters (red ovals) in (e). Scale bars, 200 nm (b,e) and 5 μm (d).

sptPALM

Synapses, including presynaptic and postsynaptic specializations are smaller than a few hundred nm and separated by only tens of nm.



E	State F1	State F2	State B2	State B1
Diff. Coef.	1.89µm²/s	0.50µm²/s	0.22µm²/s	0.09µm²/s
% of total loc.	21.47	33.51	23.03	21.97
mean time	36ms	39ms	77ms	93ms

F	State F1	State F2	State B2/B1
Diff. Coef.	1.75µm²/s	0.33µm²/s	0.11µm²/s
% of total loc.	25.5	42.7	31.8
mean time	45ms	51ms	119ms

Gomez Lamarca M. et al., Dev. Cell., 2018

CAIC

- Kevin O'Holleran
- Martin Lenz
- Ruth Sims (Institut de la Vision)
- Sohaib Abdul-Rehman
- Adam Clark
- Sarah Benaissa
- Alex Collins
- Bogdan Toader

Sainsbury Lab

- Ray Wightman

Funding

- EPSRC
- Wellcome Trust ISSF, Newton Trust

Collaborators:

- Lee group, (Dept. of Chemistry, University of Cambridge)
- Bray group, White group (PDN, University of Cambridge)

Thank you!