Efficient Large Scale Biomedical Data Analysis

Dr. Tung-Han Hsieh

Research Center for Applied Sciences,

Academia Sinica

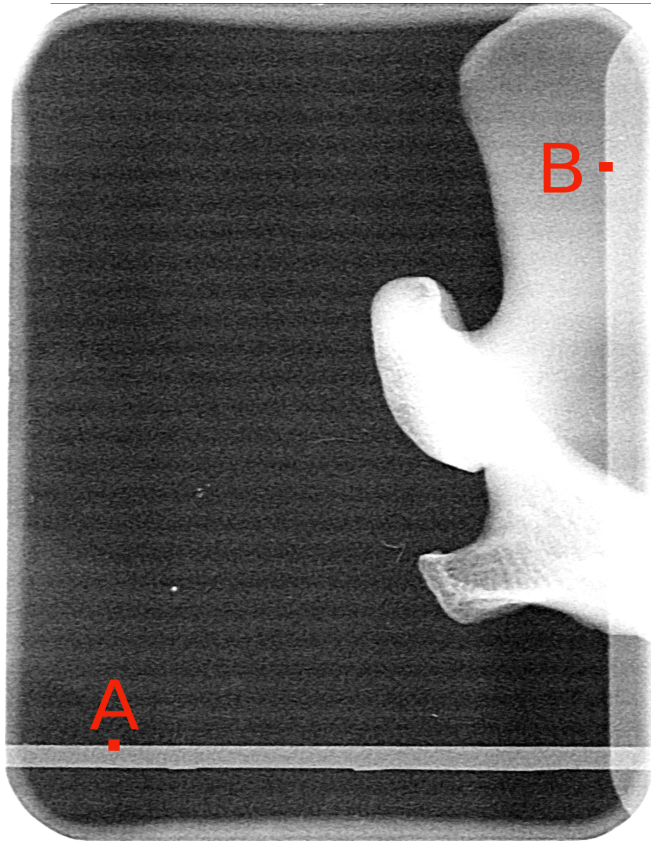
Outline

- Artificial pattern finding
  - Straight-line finding
  - Applications:
    - Image stitching
    - SPR signal analysis
- Super resolution
  - Single molecule localization (SML)
- Particle tracking
  - Improve the cmeAnalysis code
- Outlooks
  - Intelligent Image Data Processing
  - Large scale and high throughput processing.

Artificial pattern finding

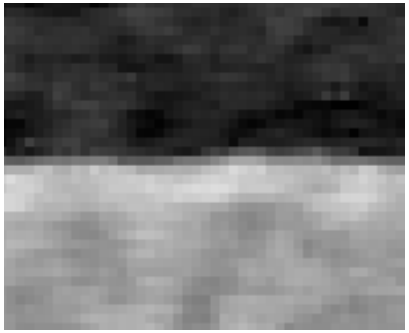
- Straight-line finding
- Applications
  - Image stitching
  - SPR signal analysis

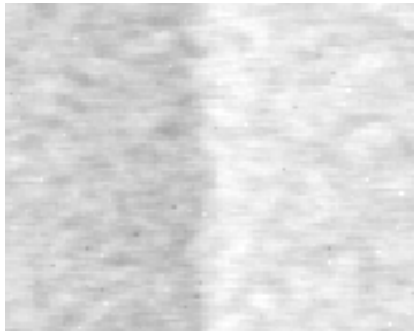
Artificial pattern finding



- Adding artificial pattern into an image is a convenient way to locate positions of image objects. However, sometimes it is non-trivial to identify these patterns.
  - Pattern intensity may be un-uniform.
  - Pattern may be overlapped by other objects.
  - Pattern boundary may not be sharp.
  - Pattern may be indistinguishable to background.

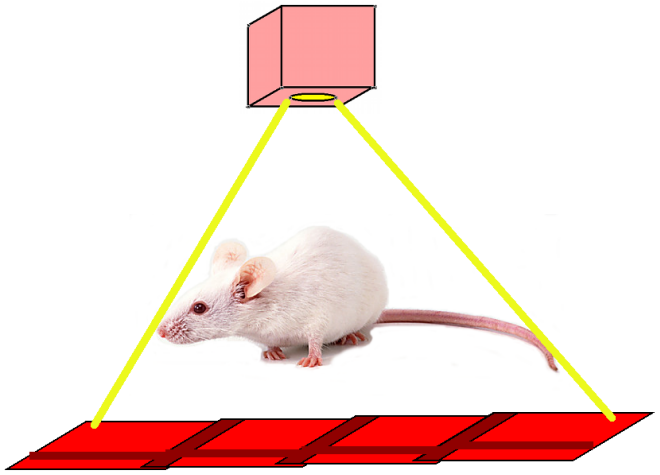
- The pattern boundary could be determined by the location of maximum difference of intensity between neighboring pixels, provided:
  - These locations form the pattern we want.
  - Statistics analysis of these locations helps us to determine the pattern correctly.

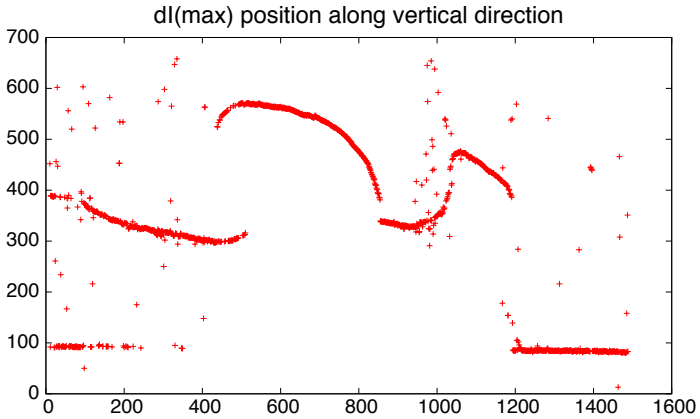
A 

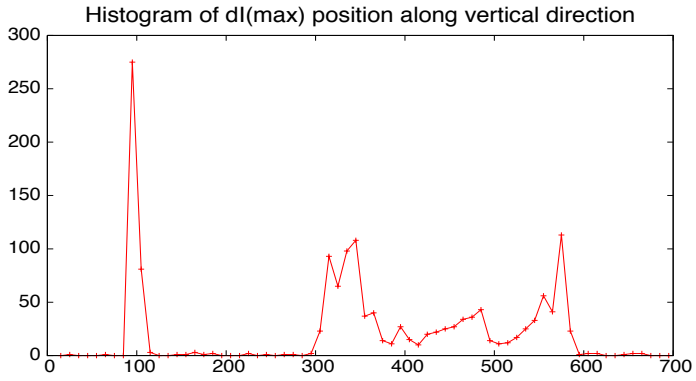
B 

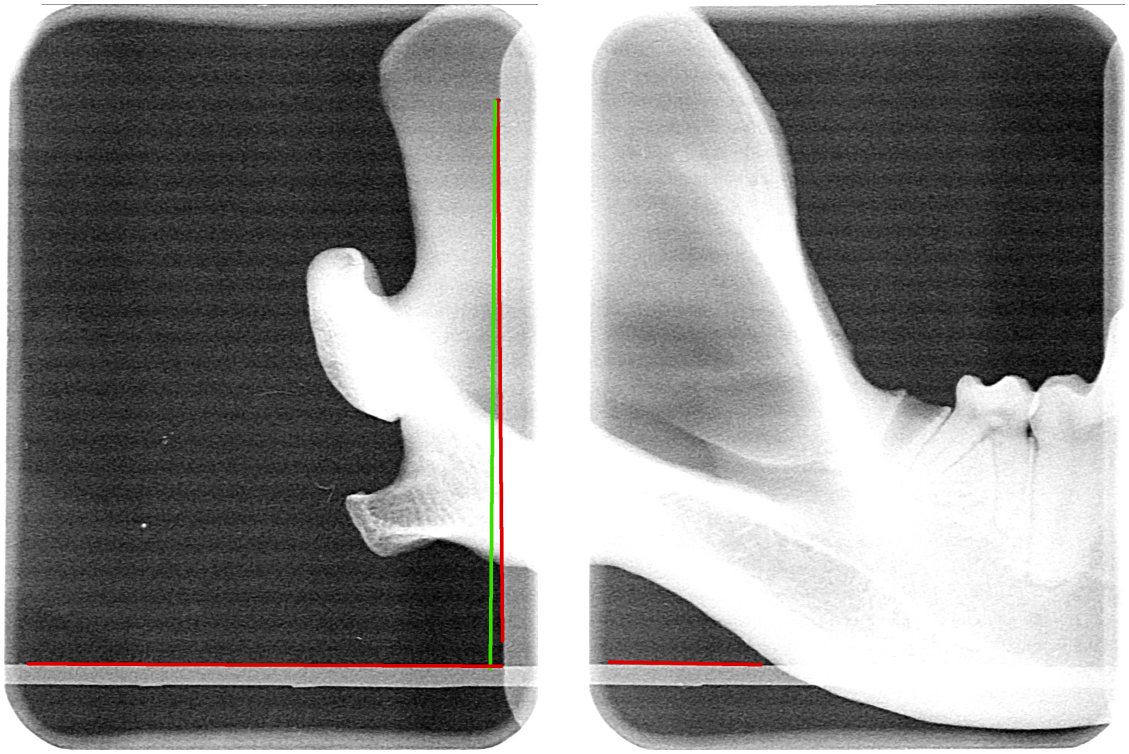
Collaborate with Jyh-Cheng Chen (MYU)

Artificial pattern finding (ex.1: Image stitching)



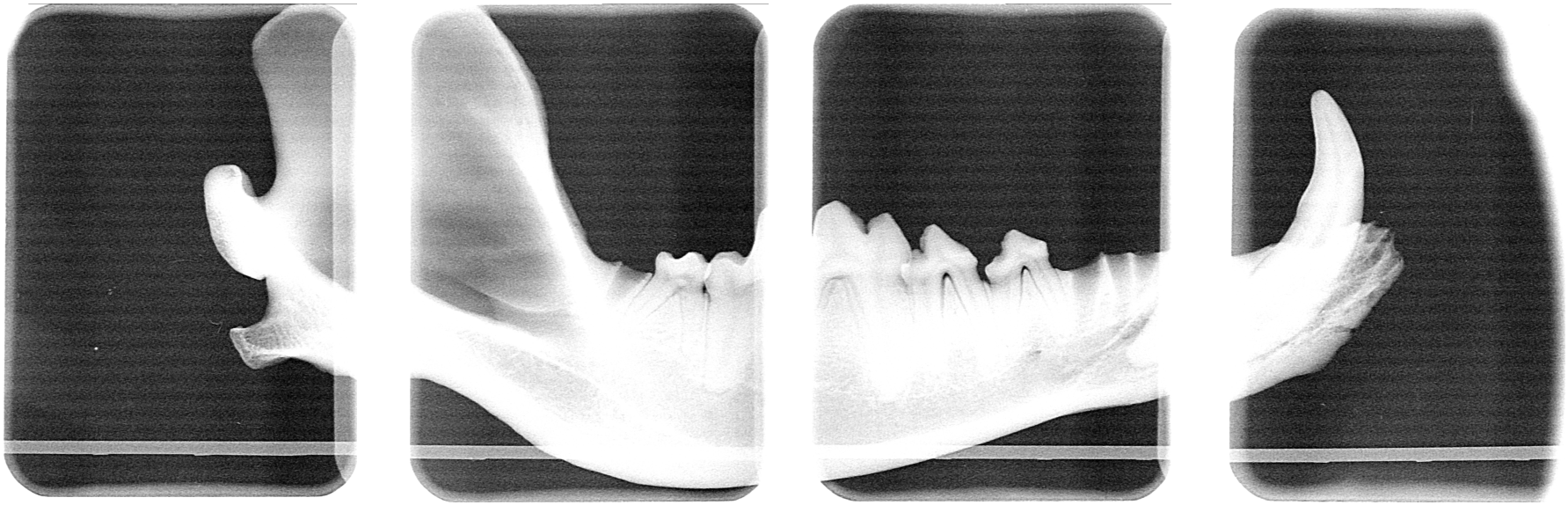


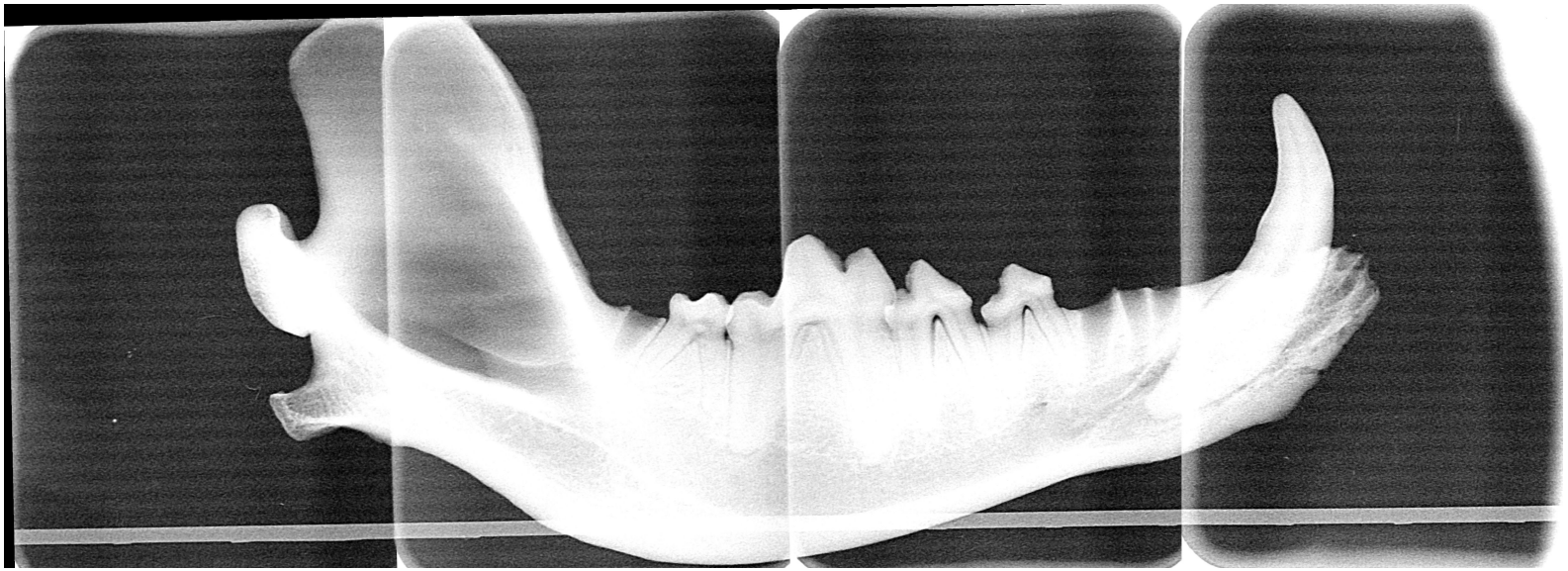




- Image stitching can be applied to take object image that is larger than the size of a single detector.
- The artificial lines are used to align the image of each part of the object for stitching.

Artificial pattern finding
(ex.1: Image stitching)



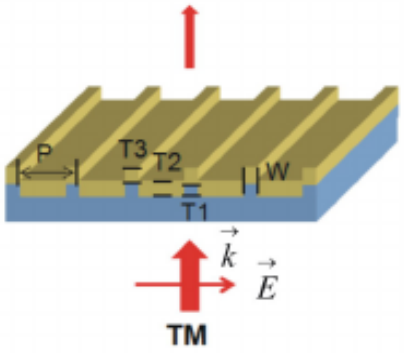


Artificial pattern finding (ex.2: SPR signal analysis)

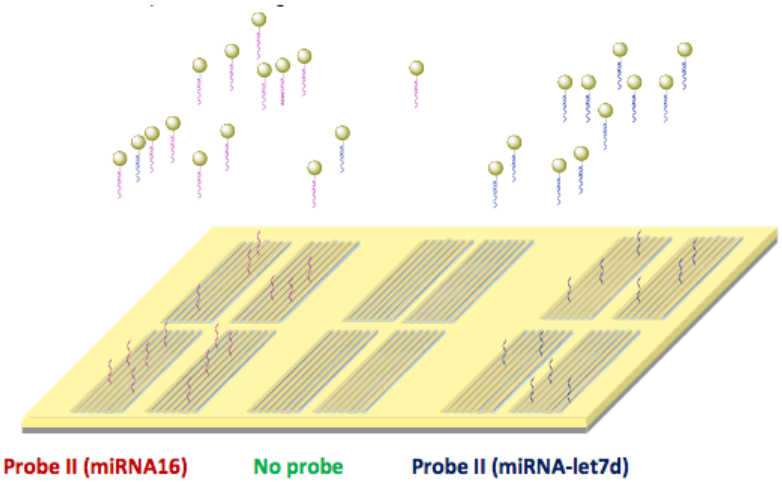
- Contrast medium induced nephropathy (CIN) (顯影劑引發的腎病) has become one of the major reasons of acute kidney injury (AKI) (急性腎損傷). Currently it cannot be diagnosed until 48 to 72 hours, which is too late for effective treatment.
- The goal:
  - To find and validate the miRNA in urine samples as the biomarkers for early stage human AKI and CIN diagnosis.
  - To develop a portable SPR system to rapidly detect the miRNA biomarkers in urine for clinical diagnosis.
- For data analysis:
  - Develop the software for SPR signal analysis.
  - Develop the integrated software to operate the portable SPR system.

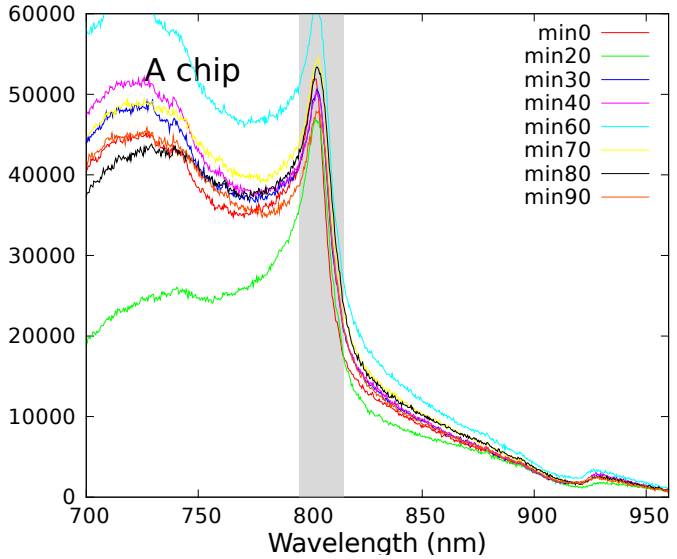
Collaborators: Ji-Yen Cheng (AS), Pei-Kuen Wei (AS), Heng Lin (TMU), and Hsi-Hsien Chen (TMU)

Artificial pattern finding (ex.2: SPR signal analysis)



K.L.Lee, *et al*, Scientific Report (2015).



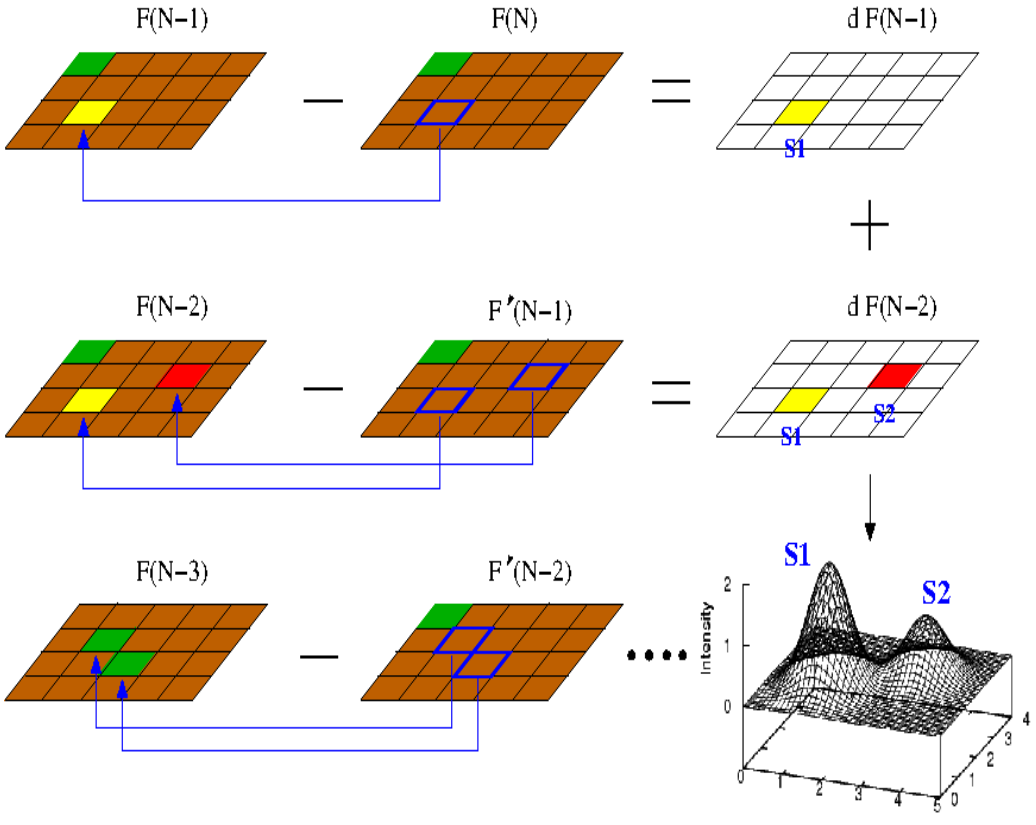


- The probes for biomarkers are attached to microfluid chips which embed with the sensitive nanoslit SPR chips.
- The presence of target biomarkers is revealed in the shift of SPR peak in the spectrum.
- The SPR signal analysis relies on examining the SPR shift w.r.t. time.

Super Resolution

(Single Molecule Localization)

Super resolution
(Single Molecule Localization Algorithm)

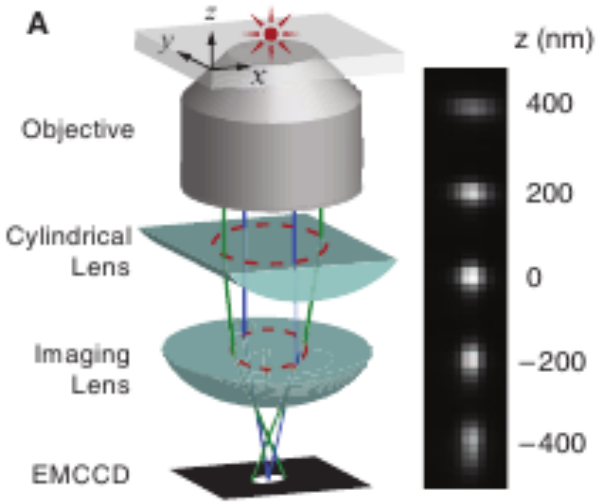


$$I(x,y) = A\exp\left\{-\frac{2(x-x_0)^2}{w_x^2} - \frac{2(y-y_0)^2}{w_y^2}\right\} + I_B$$

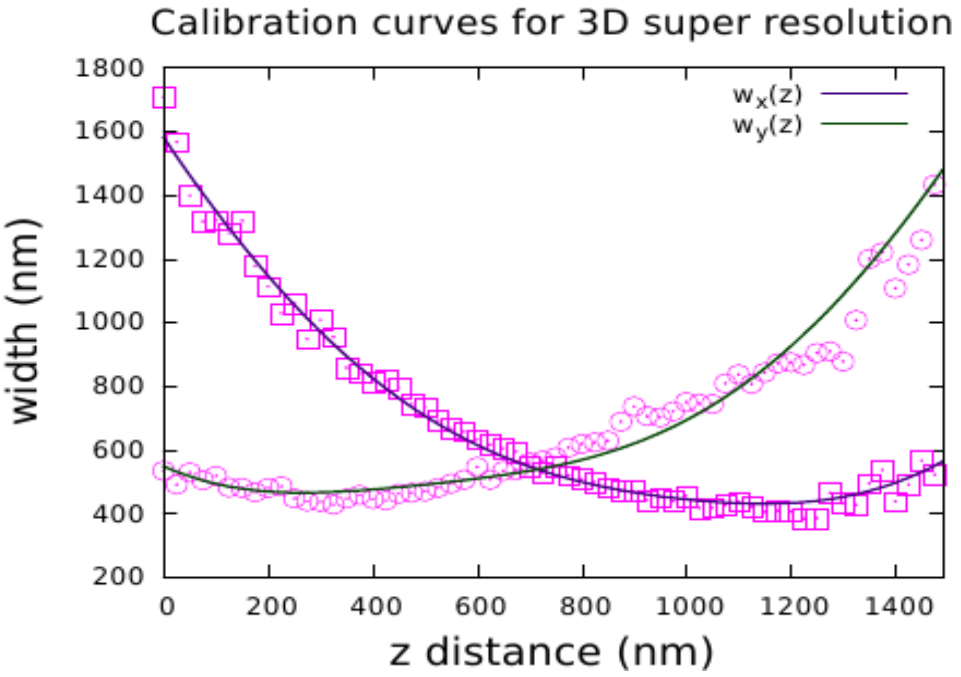
- In PALM & STORM super resolution imaging, the target protein molecule is attached with a fluorescence molecule.
- The fluorescence molecule is activated by a pulse of laser, and deactivated by photon emission (PALM) or another pulse of laser (STORM).
- Taking successive images of the fluorescent sample in activated – deactivated cycles, we can precisely determine molecules' location by SML algorithm.

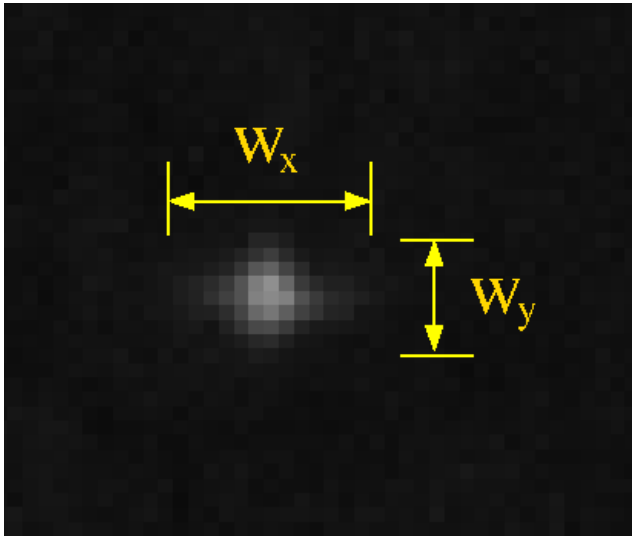
Collaboration with Peilin Chen, AS.

Super resolution (3D)



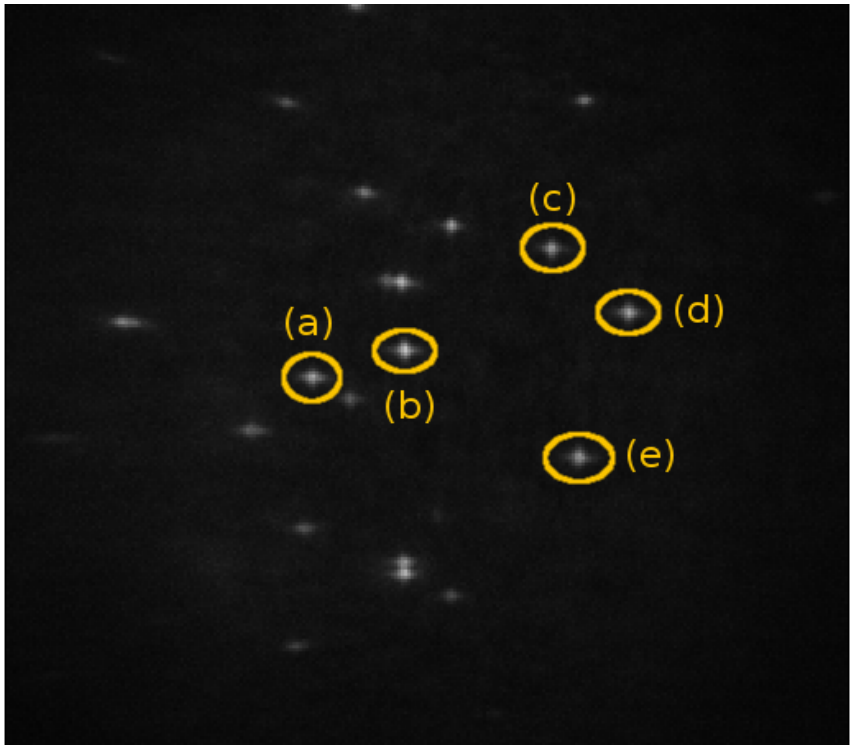
Ref: Huang, et al, Science (2008)

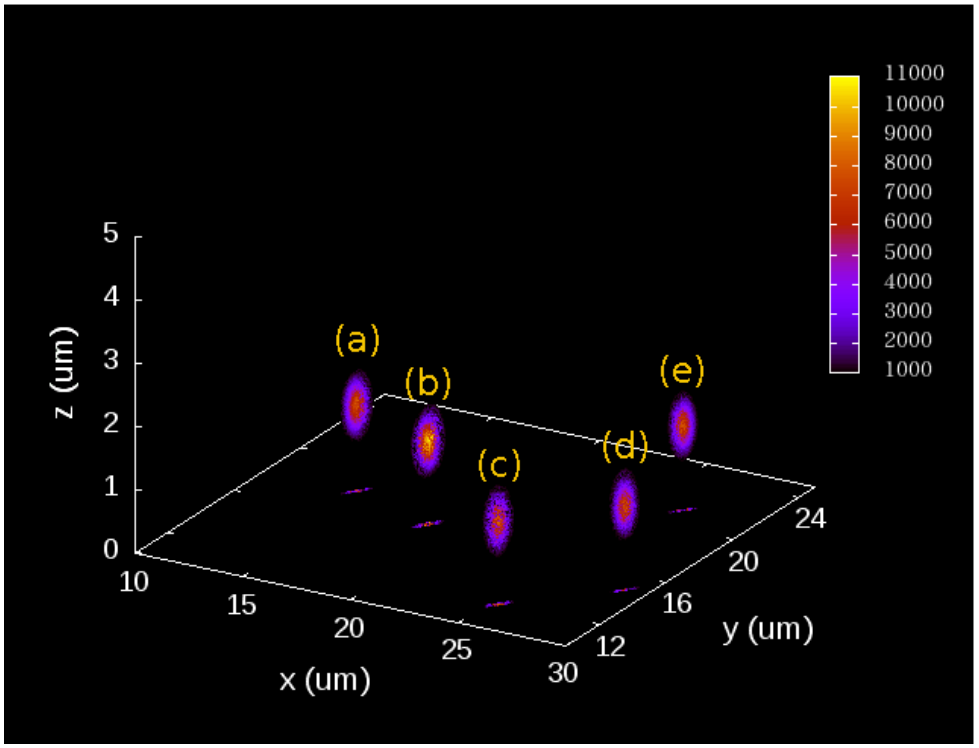


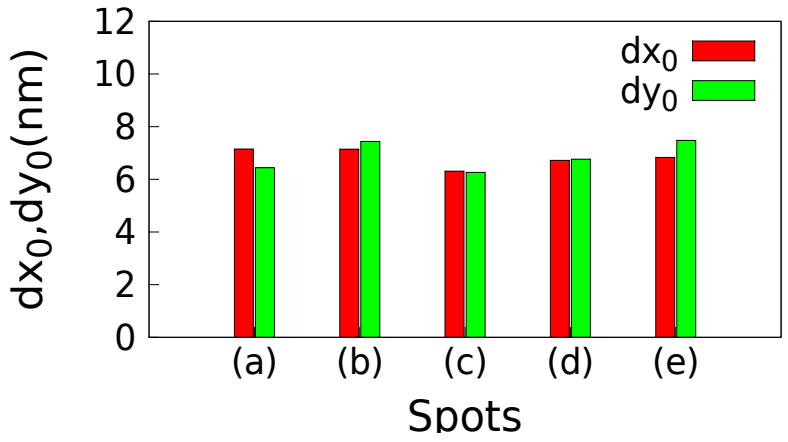


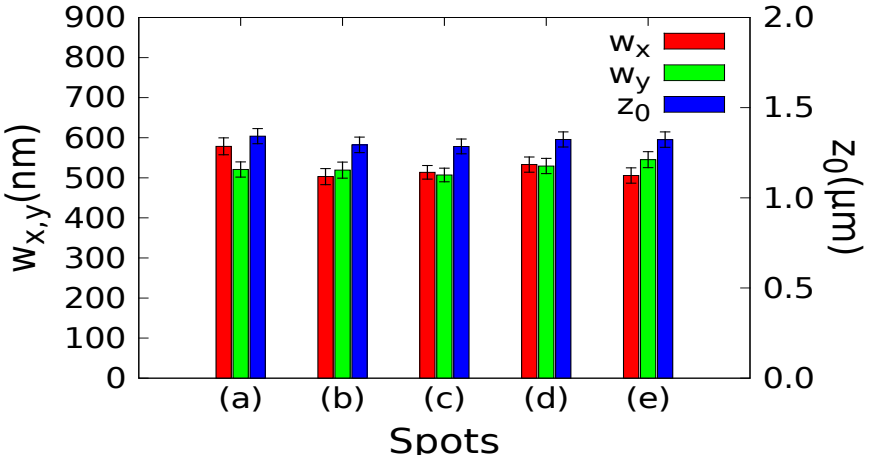
A set of cylindrical lens is introduced in the light path of image to distort the spots from circle to ellipse. The widths w_x & w_y depends on the vertical position of the molecule. As a result, we can use the calibration curve to determine the *z* coordinate for the given w_x & w_y.

Super resolution (testing results)









Particle Tracking

(Improve the cmeAnalysis code)

Particle tracking

- The Granulocyte-Macrophage colony stimulating factor Receptor (GMR) (粒細胞 – 巨噬細胞製造之激發因子受體), which is composed of α and β subunit, is important to regulating cell proliferation (增殖) and differentiation (分化) in each lineages (譜系).
- The goal: To demonstrate the endocytosis (內吞作用) and activation machinery of GM-CSF (Colony Stimulating Factor) receptor by two color-tracking analysis of TIRF images for GMRα & GMRβ.
- For data analysis, we developed the particle tracking tool to track the movement of two types of particles through different channels and reconstruct the co-localized tracks.
- This work is based on the cmeAnalysis code developed by *Danuser Lab* at Harvard.

Collaborate with Jeffrey j.Y. Yen (AS) & Fan-Ching Chien (NCU)

Particle tracking (cmeAnalysis)

Danuser Lab - Iceweasel

Danuser Lab

lccb.hms.harvard.edu/software.html

Search

Most Visited | Getting Started | Latest Headlines

danuser lab

harvard medical school :: department of cell biology :: laboratory for computational cell biology

home | research | people | publications | software | contact | intranet

software

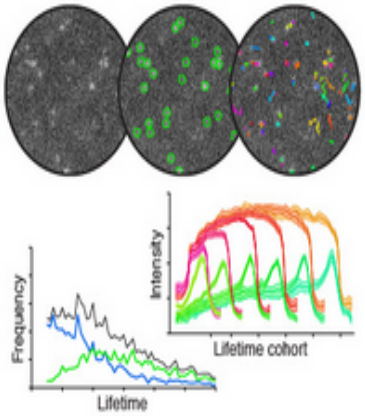
cmeAnalysis

cmeAnalysis is a Matlab software package for the quantification of clathrin-coated pit dynamics from fluorescence time-lapse data. The functionalities provided include highly sensitive detection, tracking (based on u-track), master/slave detection for multi-channel data, intensity-based classification of coated structures, and lifetime analysis. Additionally, the package includes a graphical user interface for inspection of analysis results from individual movies.
The download includes a user guide, and a test data set is available separately.

- Download (3.5 Mb)
- Test data set (300 Mb)

Release notes: version 1.03 includes improvements to graphical and console output.

For more information, please see Aguet et al., *Dev. Cell* 26(3), pp. 279-291, 2013.



http://lccb.hms.harvard.edu/software.html

Particle tracking
(Original paper of cmeAnalysis)

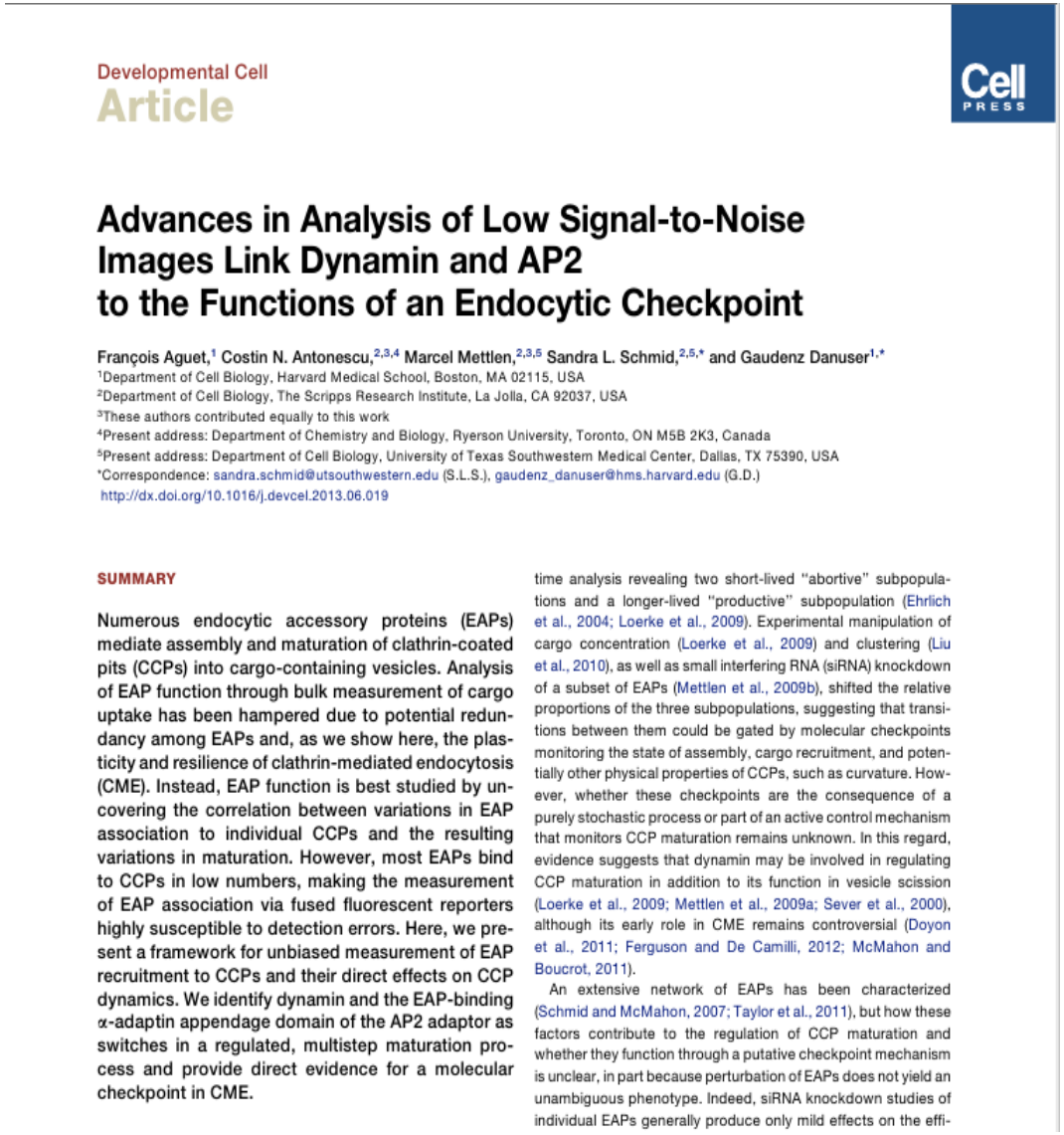
F.Aguet, *et al*, Developmental Cell 26, 279-291 (2013)

Advantages

- Suitable for us
- Widely used
- Source code available

Problem

- Two-Channel tracking: EGFP as master channel, and RFP as slave channel, but not independent channels.
- UI is not flexible and not user friendly.
- Tracking data is not available.



Developmental Cell
Article

Cell PRESS

Advances in Analysis of Low Signal-to-Noise Images Link Dynamin and AP2 to the Functions of an Endocytic Checkpoint

François Aguet,[1] Costin N. Antonescu,[2,3,4] Marcel Mettlen,[2,3,5] Sandra L. Schmid,[2,5,*] and Gaudenz Danuser[1,*]
[1]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Cell Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
[3]These authors contributed equally to this work
[4]Present address: Department of Chemistry and Biology, Ryerson University, Toronto, ON M5B 2K3, Canada
[5]Present address: Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
*Correspondence: sandra.schmid@utsouthwestern.edu (S.L.S.), gaudenz_danuser@hms.harvard.edu (G.D.)
http://dx.doi.org/10.1016/j.devcel.2013.06.019

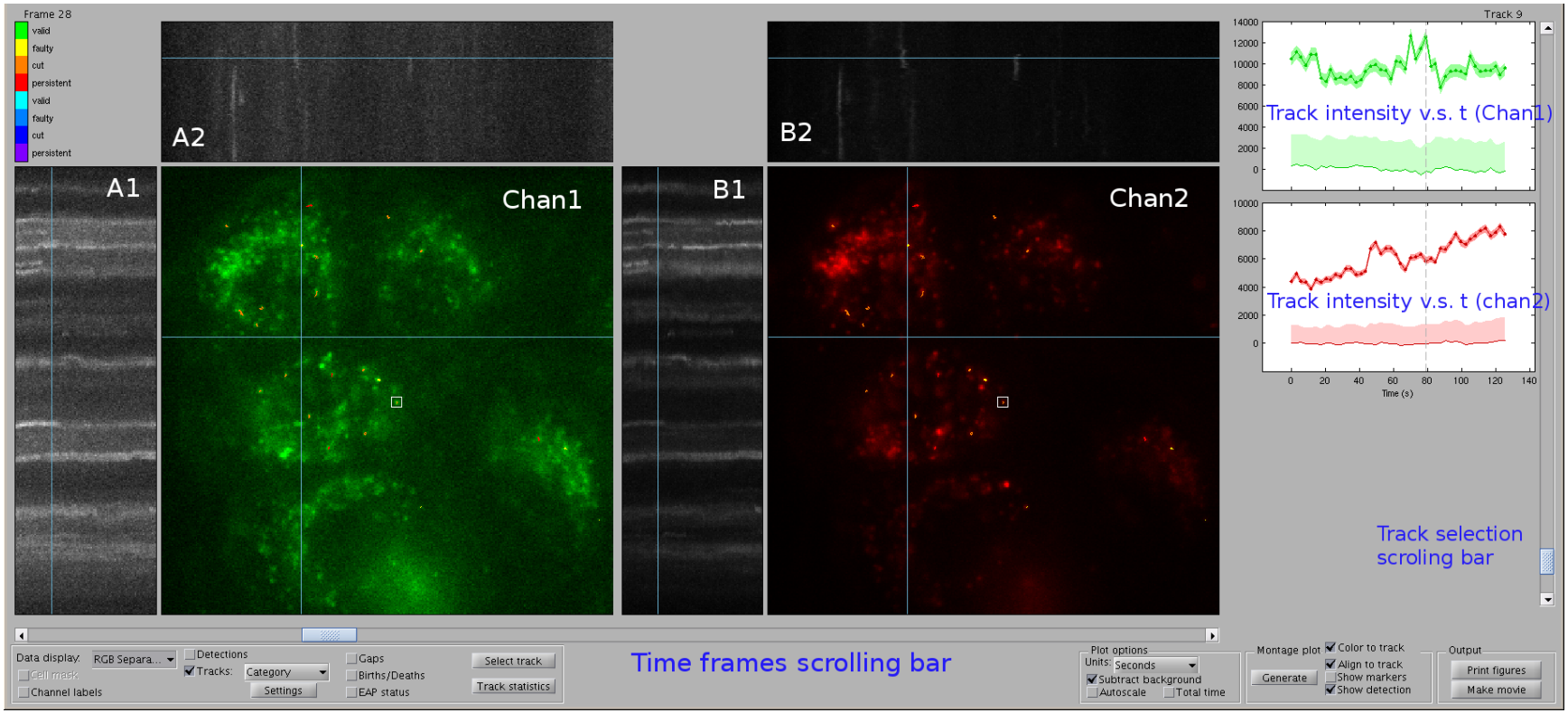
SUMMARY

Numerous endocytic accessory proteins (EAPs) mediate assembly and maturation of clathrin-coated pits (CCPs) into cargo-containing vesicles. Analysis of EAP function through bulk measurement of cargo uptake has been hampered due to potential redundancy among EAPs and, as we show here, the plasticity and resilience of clathrin-mediated endocytosis (CME). Instead, EAP function is best studied by uncovering the correlation between variations in EAP association to individual CCPs and the resulting variations in maturation. However, most EAPs bind to CCPs in low numbers, making the measurement of EAP association via fused fluorescent reporters highly susceptible to detection errors. Here, we present a framework for unbiased measurement of EAP recruitment to CCPs and their direct effects on CCP dynamics. We identify dynamin and the EAP-binding α-adaptin appendage domain of the AP2 adaptor as switches in a regulated, multistep maturation process and provide direct evidence for a molecular checkpoint in CME.

time analysis revealing two short-lived "abortive" subpopulations and a longer-lived "productive" subpopulation (Ehrlich et al., 2004; Loerke et al., 2009). Experimental manipulation of cargo concentration (Loerke et al., 2009) and clustering (Liu et al., 2010), as well as small interfering RNA (siRNA) knockdown of a subset of EAPs (Mettlen et al., 2009b), shifted the relative proportions of the three subpopulations, suggesting that transitions between them could be gated by molecular checkpoints monitoring the state of assembly, cargo recruitment, and potentially other physical properties of CCPs, such as curvature. However, whether these checkpoints are the consequence of a purely stochastic process or part of an active control mechanism that monitors CCP maturation remains unknown. In this regard, evidence suggests that dynamin may be involved in regulating CCP maturation in addition to its function in vesicle scission (Loerke et al., 2009; Mettlen et al., 2009a; Sever et al., 2000), although its early role in CME remains controversial (Doyon et al., 2011; Ferguson and De Camilli, 2012; McMahon and Boucrot, 2011).

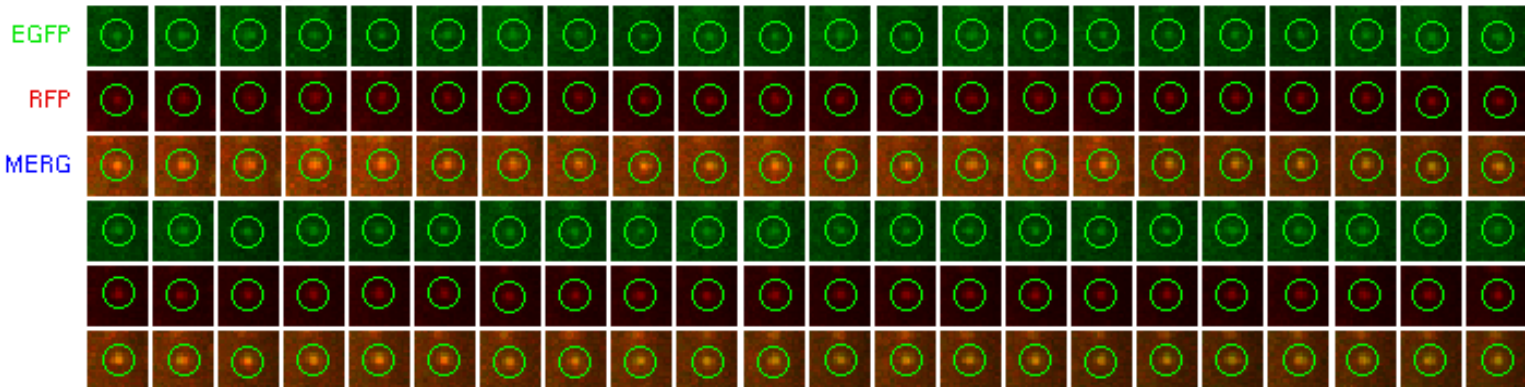
An extensive network of EAPs has been characterized (Schmid and McMahon, 2007; Taylor et al., 2011), but how these factors contribute to the regulation of CCP maturation and whether they function through a putative checkpoint mechanism is unclear, in part because perturbation of EAPs does not yield an unambiguous phenotype. Indeed, siRNA knockdown studies of individual EAPs generally produce only mild effects on the effi-

Particle tracking (cmeAnalysis)

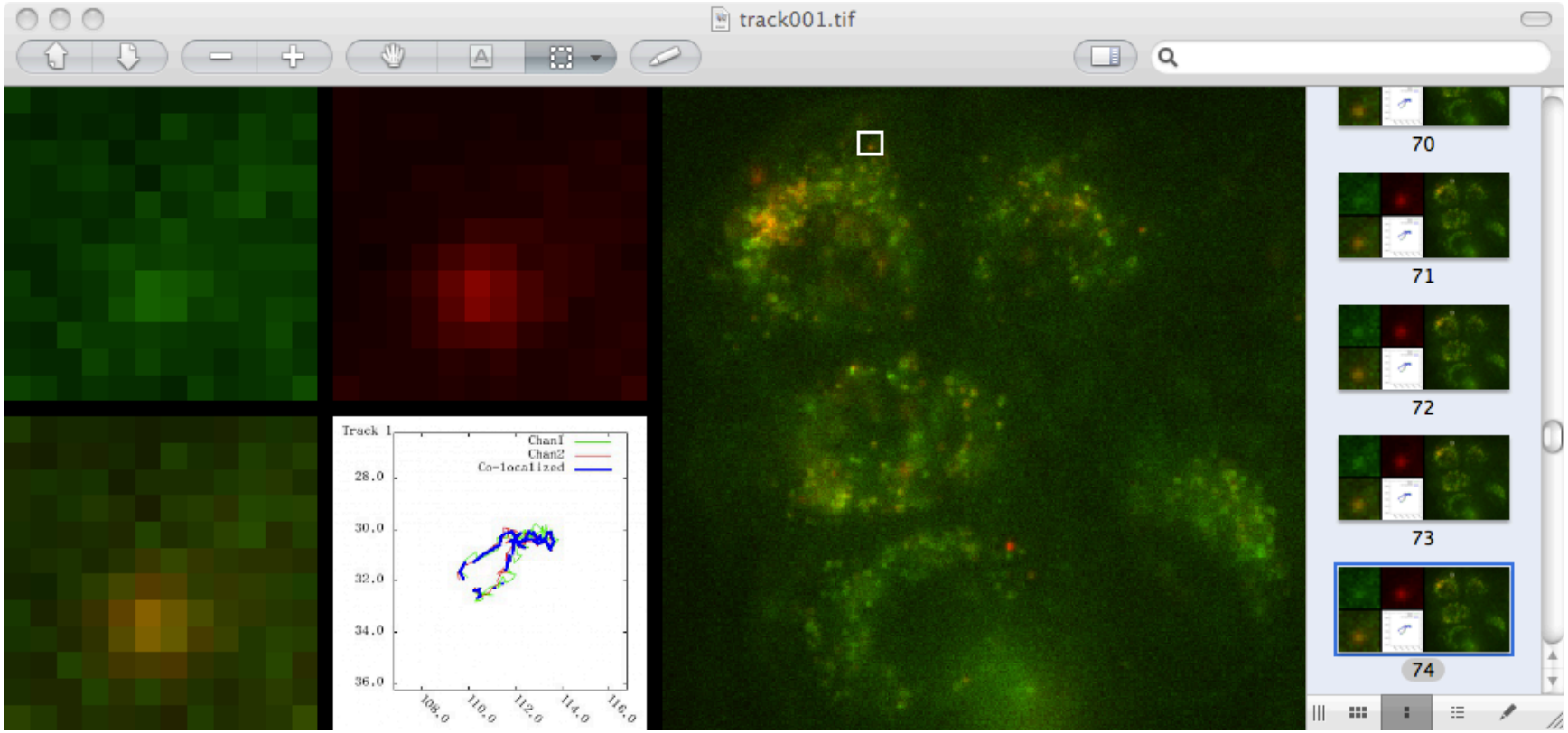


GMRα

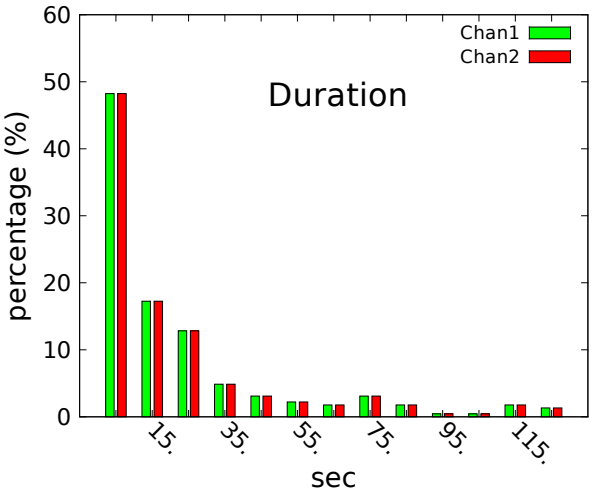
GMRβ

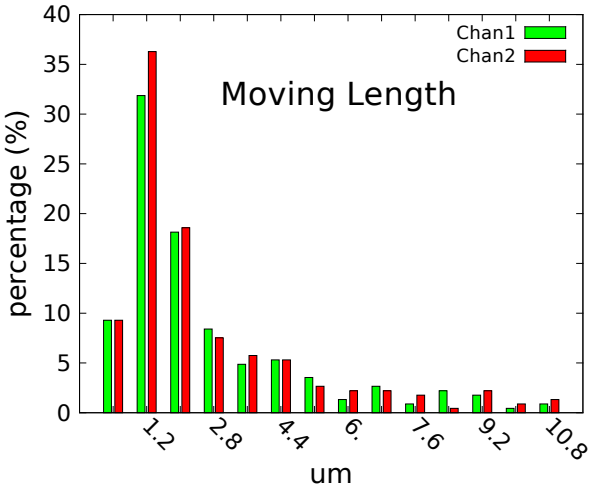


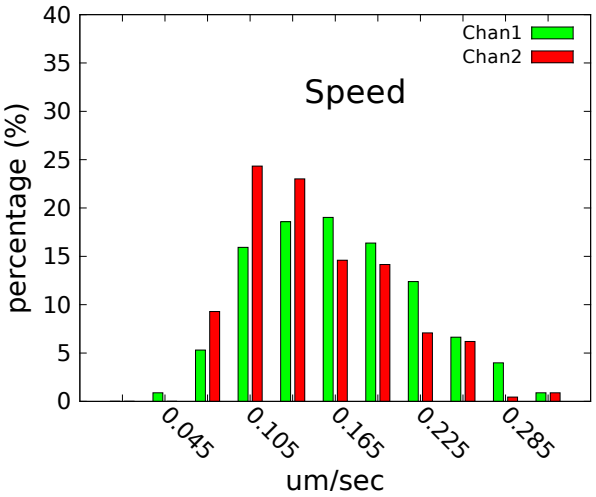
Particle tracking (detailed track viewer)



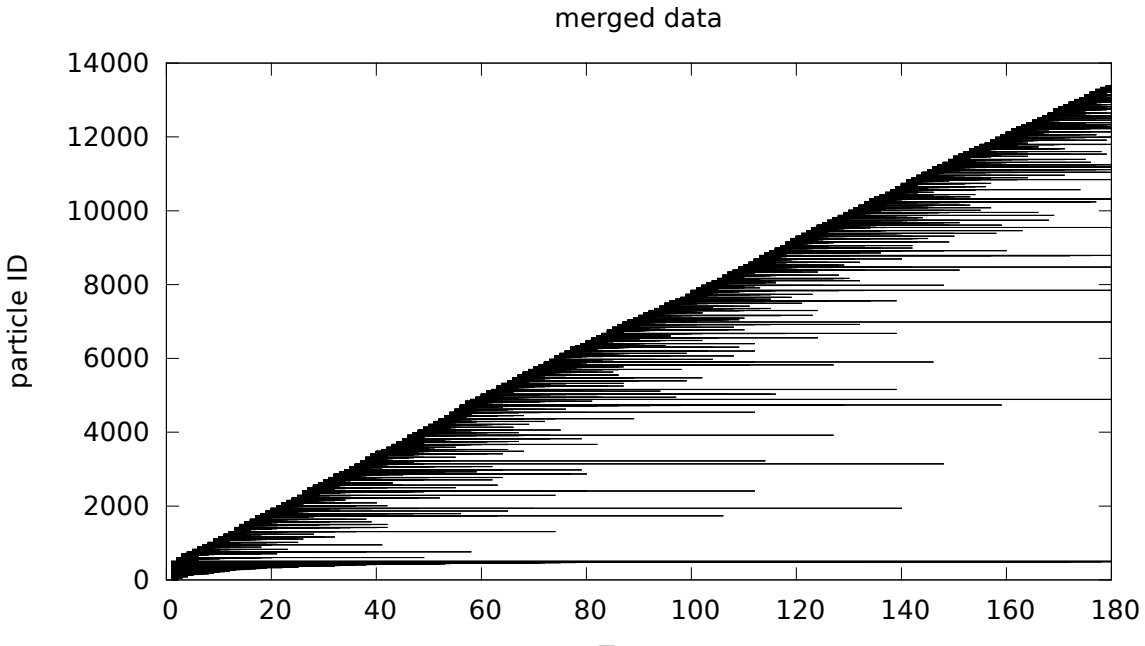
Histogram of particle movement







Distribution of Particle Durations



Outlooks

- Intelligent image data processing
- Large scale and high throughput processing

Outlooks

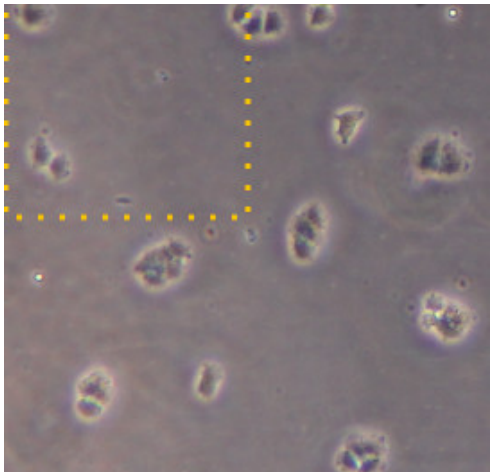
We are facing the challenges in biomedical image processing tasks: *preciseness* and *performance*.

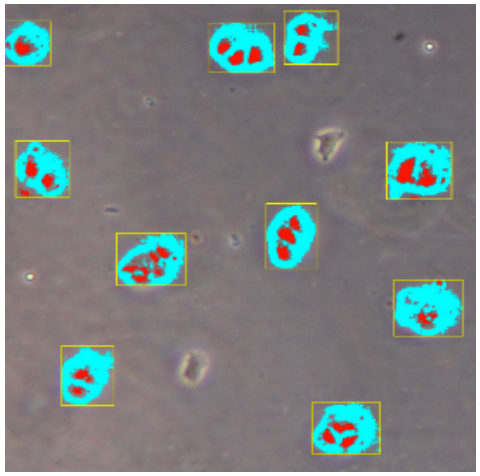
To tackle these challenges:

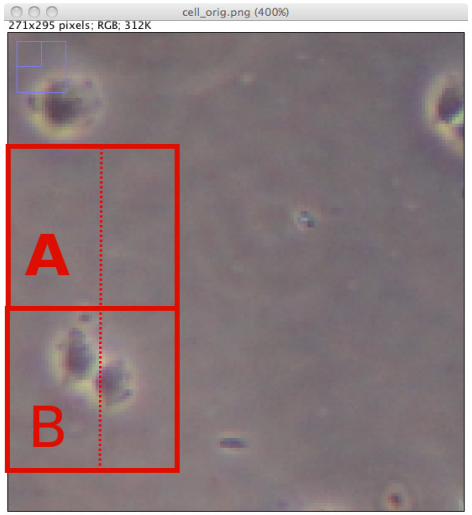
- Intelligent image data processing:
  - Testing and making decision.
  - Machine learning.
- Large scale and high throughput processing:
  - Good performance to analyze large amount of data
  - Good user interface to navigate the data and the analysis results.

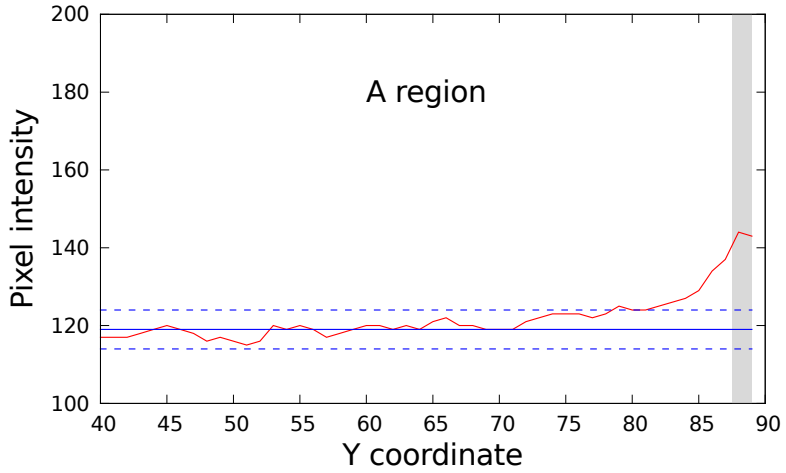
Outlooks

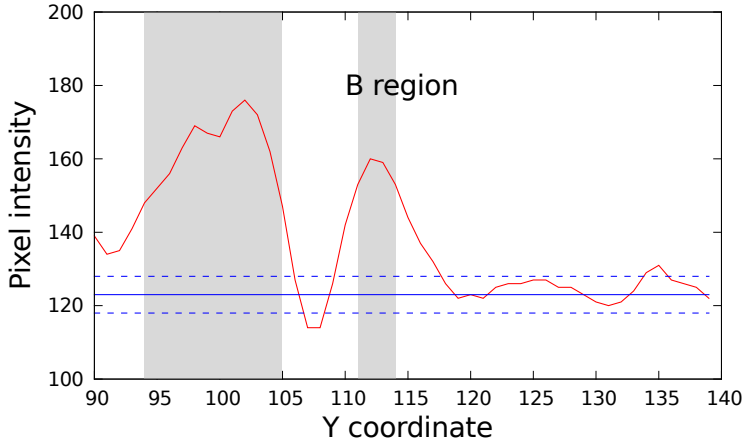
(Problem of preciseness: Cell image analysis)







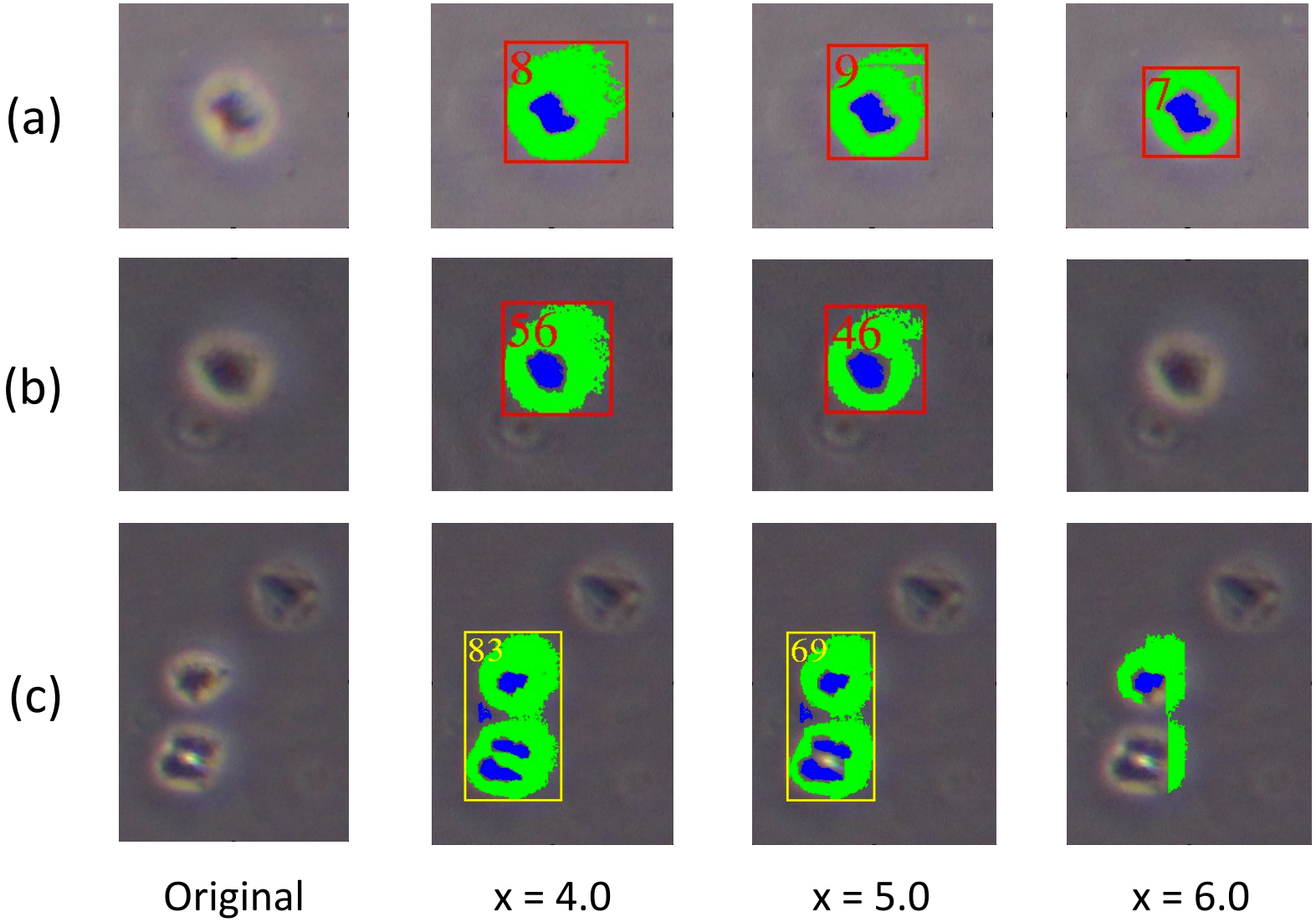




Collaboration with Ji-Yen Cheng (AS)

Outlooks

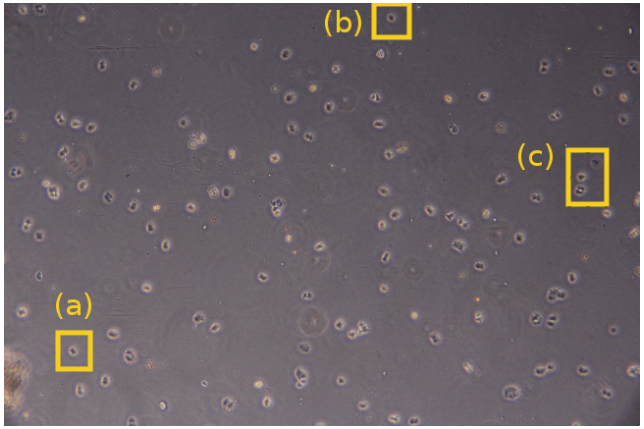
(Problem of preciseness: Cell image analysis)



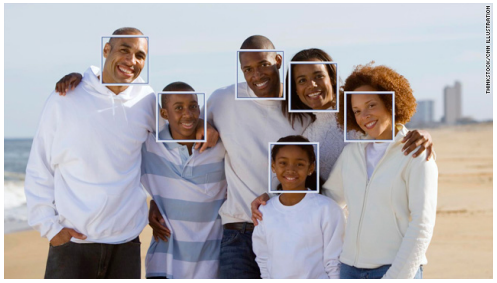
Outlooks

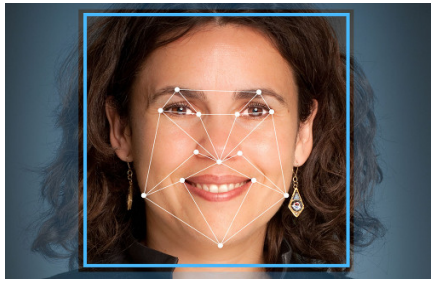
(Cell image analysis: the proposed ideas)

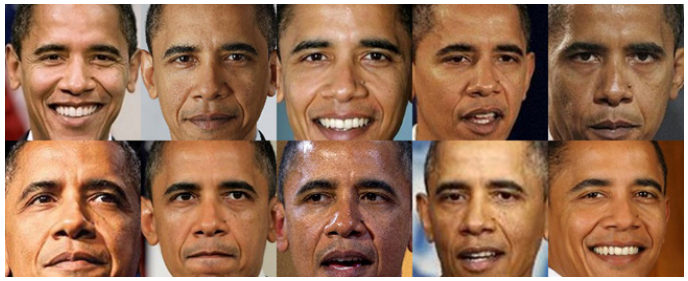
- Testing and making decision:
  - The cell image should be examined in different scales, to capture the cell locations and filter out the background.
  - The cell boundary should be tested with different thresholds.
  - Making scores for each test, and finally making the decision.



- Machine learning:
  - Identify the pattern features of a given image.
  - Fit the result to the model derived from the learning process.

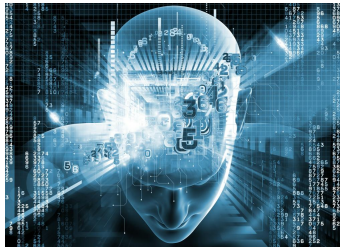




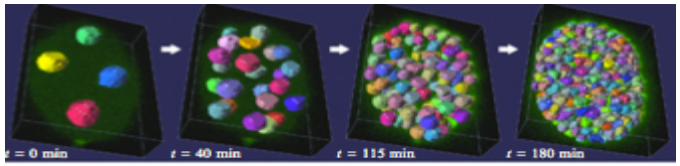


Outlooks





- Biomedical research heavily relies on bioimage analysis.
- Particle tracking with super resolution:
  - Track huge amount of individual particles in nano-scale.
  - At each sampled time, >= 10000 image frames are required for SML algorithm.



- Cell tracking:
  - Has to analysis and incorporate features of each cell for tracking.
  - Cell lineage reconstruction.
- More importantly, how to process and access the large amount of data, or *Big Data* ?
  - Powerful computing system (parallel / GPU / Cloud / AI computing …)
  - Transfer, store, annotate, share, search, and derive new biomedical knowledge.
  - Automatic data processing and analysis are required.
  - We also need a well designed user interface and workflow to access and process data.

Thanks for your attention.