Quantification of Cytoskeletal Protein Localization from High-Content Images

Shiwen ZHU, Paul MATSUDAIRA,

Roy WELSCH, Jagath C. RAJAPAKSE

Singapore-MIT Alliance

Cytoskeleton

(www.wikipedia.org)

- **Cellular "skeleton" or "scaffold"**
- **Function as dynamic structure**
    - Maintaining cell shape
    - Protecting cells
    - Supporting cellular motion
    - Intracellular transporting
    - Cell division
- **Cytoskeletal proteins generate harmonious responses to the coordinated efforts of cellular networks.**

- **However, very little is known about the coordinated system of these proteins**
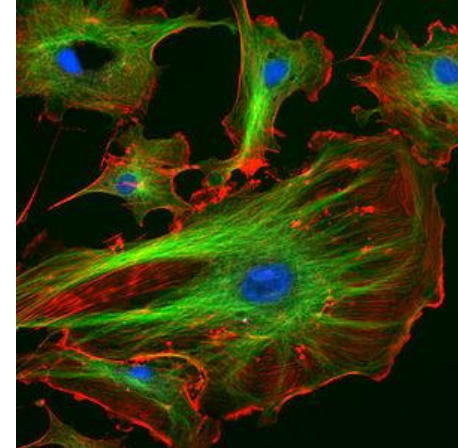
Colocalization

- **Colocalization refers to the degree of overlap between the targets indicated by two different fluorescent labels**

| Global statistic approaches	Authors and years	Object-based approaches	Authors and years
Pearson coefficient	E. Manders, et al., 1992	Centroid position or Intensity centers	Y. Boutte, et al., 2006
Overlap coefficient	E. Manders, et al., 1992		
Statisitcal significance	S.V. Costes, et al., 2004	Intensity correlation coefficient	F. Jaskolski, et al., 2005
Intensity correlation	Q. Li, et al., 2004		

Colocalization

$$Coloc(c, p) = \frac{\sum_{x \in W_p} f_\mathrm{p}(x) \cdot \tilde{f}_c(x)}{\sum_{x \in W_p} f_\mathrm{p}(x)}$$

- **Where W_p represents the set of pixels in the region occupied by the protein p.**
- $f_\mathrm{p}(x)$ **is the intensity distribution of the GFP channel highlighting protein p.**
- $\tilde{f}_c(x)$ **is the intensity transformation of the compartment c.**
- $c = \{nucleus, cytoplasm, actin, membrane, cytosol\}$

Framework
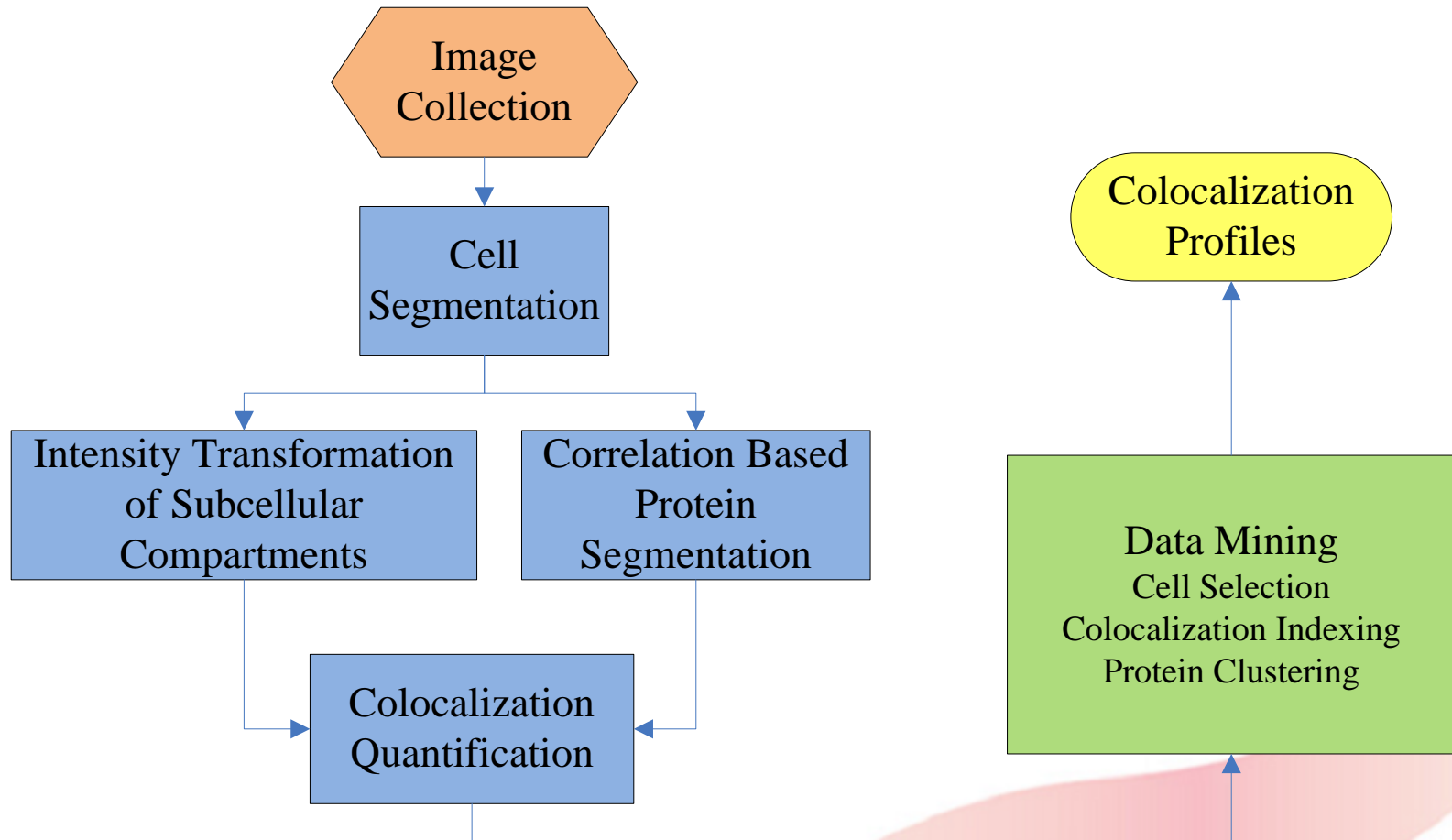
Cell segmentation

- **Region growing method with Multi-resolution segmentation**

Algorithm 1: Region growing method with multi-resolution segmentation objects

BEGIN:

$$L_n = \{l : g_n(l) \geq t_n, l \in \Gamma\}$$

$$L_a = \{l : g_a(l) \geq t_a, l \notin L_n, l \in \Gamma\}$$

$$L_c = L_n$$

$$t = \max_l \{g_a(l) + g_n(l) : l \in \Gamma\}$$

$$k = 0$$

WHILE $t \geq 0$

$$L' = \{l : (g_n(l) + g_a(l)) \geq t, l \notin L_n\}$$

$$L'' = \{l : |U(l)| > 1, l \in L'\}, \quad U(l) = \{l' : l' \in L_c, l' \in N(l), l \neq l'\}$$

$$O_l = O_l \cup O_{l'}; \ l \in L_c, l' \in \{L' \cap \overline{L''} \cap N(l)\}$$

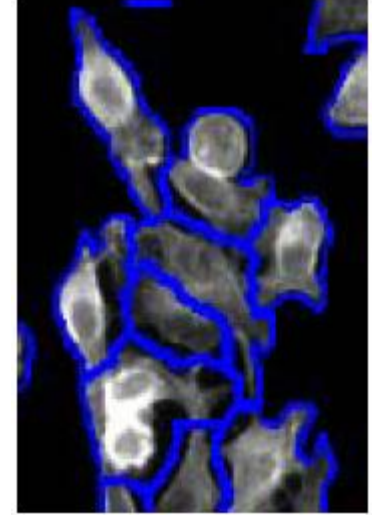$$O_n = O_l \cup O_n, \ l \in L'', \ n \in \{L_c \cap N(l)\}$$
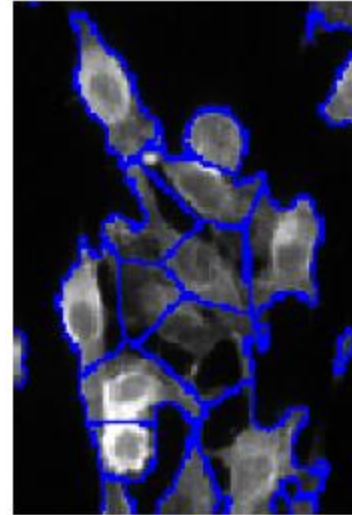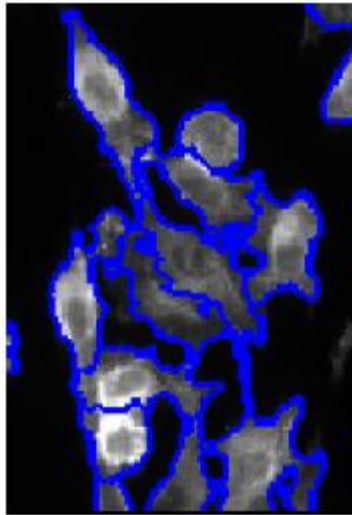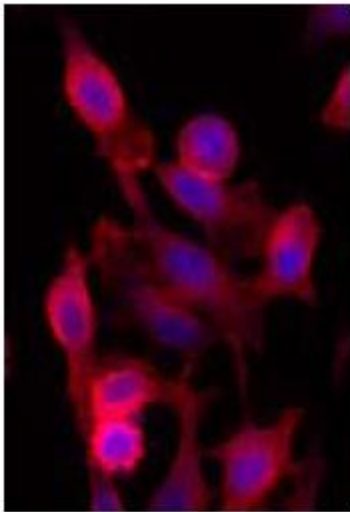
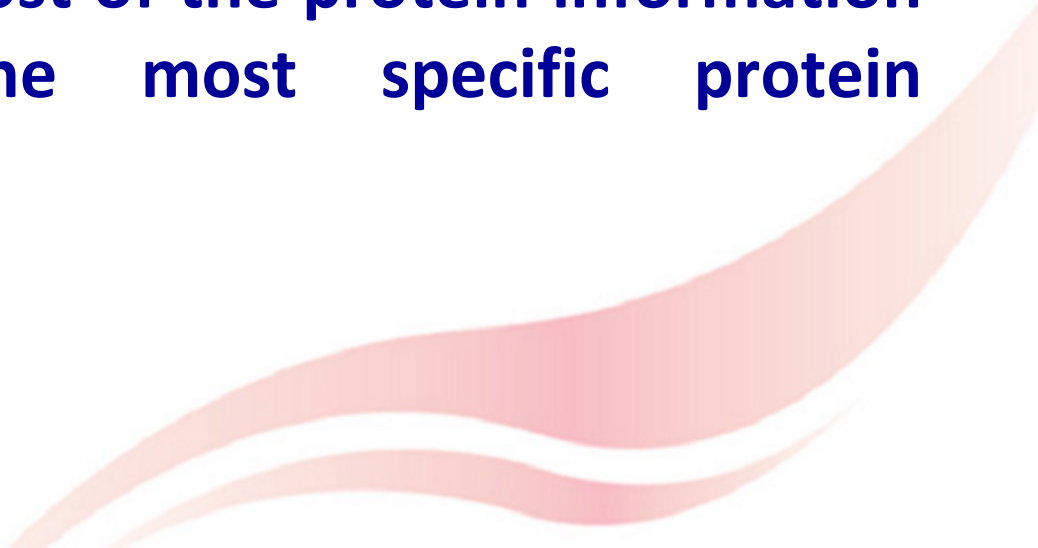$$t = t - \Delta$$

ENDWHILE

END
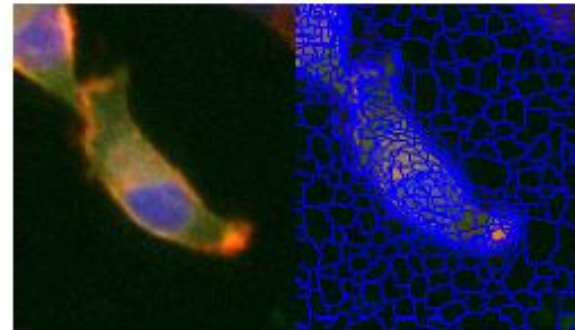
Figure 14 Example of multi-resolution segmentation

6
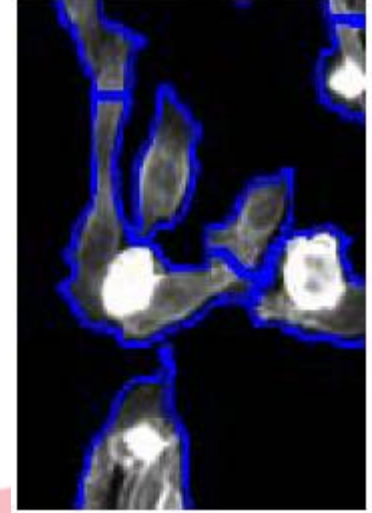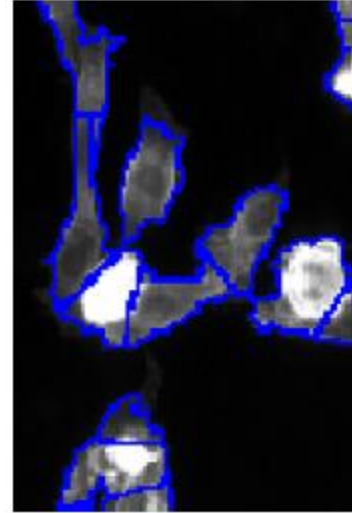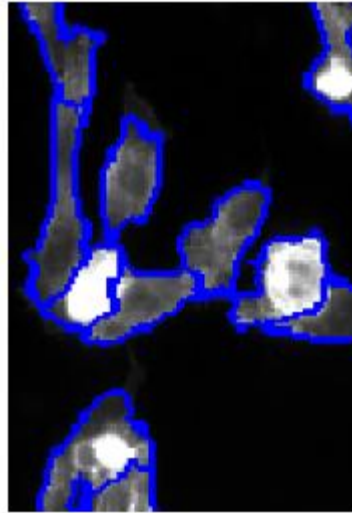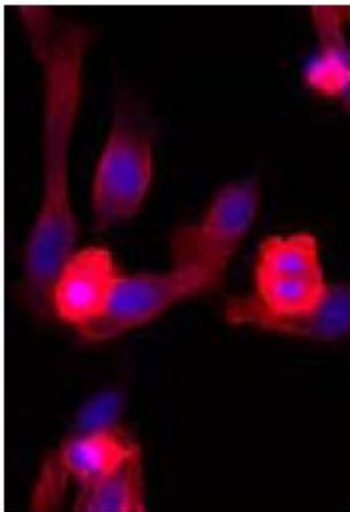
Raw images | Region growing with multiresolution segmentation | (J. Cheng, et al., 2009) | (W. Yu, et al., 2008)

7

Protein Segmentation

- **GFP intensities vary in different cells and different images.**

- **Protein segmentation is applied to set the appropriate GFP threshold maintaining the balance between capturing most of the protein information and highlighting the most specific protein information.**
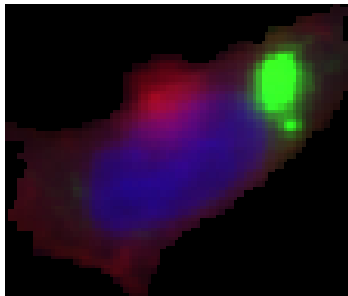
Protein Segmentation
# Thresholding based on correlation measurement

- **Find the Threshold:**
  - Use each GFP intensity level as threshold T
    - Pixels with intensities greater than T – selected pixels
    - Pixels with intensities smaller than T – unselected pixels
  - Min correlation on unselected pixels – Tmin
  - Max correlation on selected pixels – Tmax
  - Mean of Tmin and Tmax

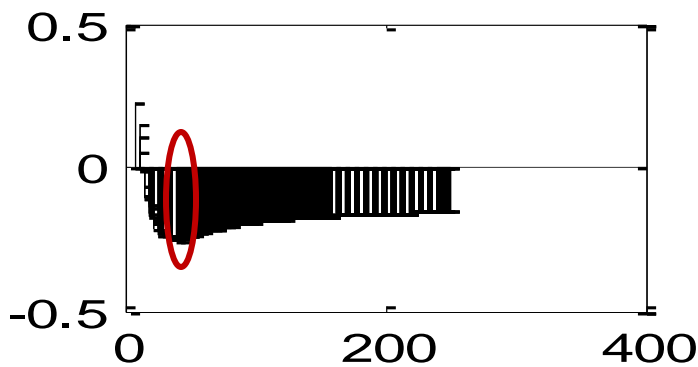# Protein Segmentation
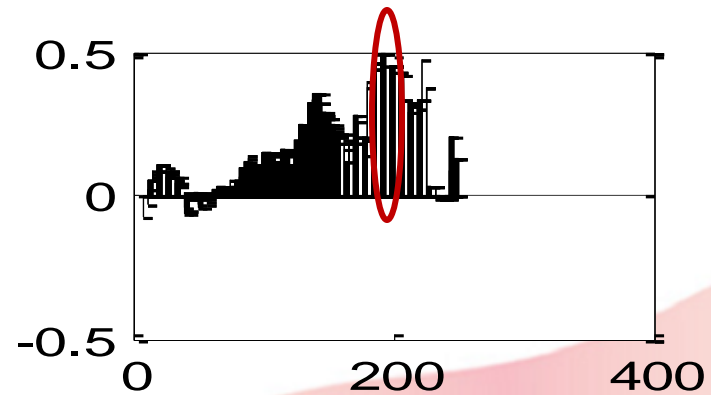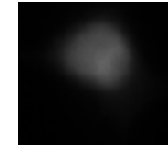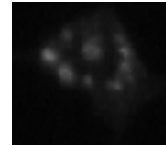# Thresholding based on correlation measurement

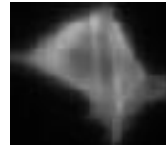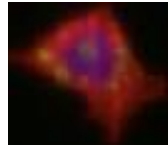Cell
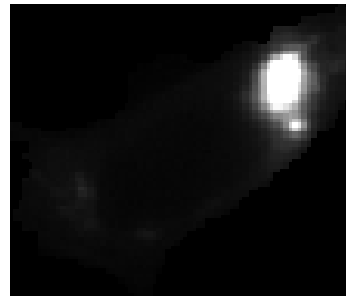
GFP

Protein
Segmentation

ICQ on unselected pixels

ICQ on selected pixels

Intensity Transformation of Subcellular Compartments- Cell Sample

Coronin 2B Protein

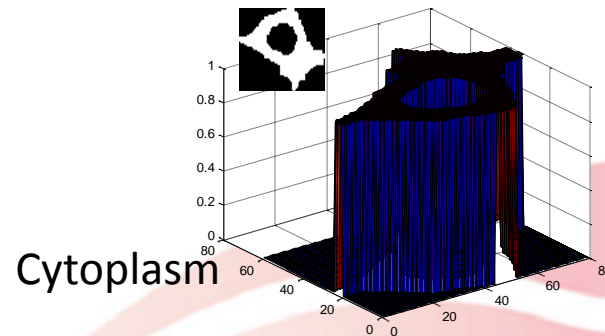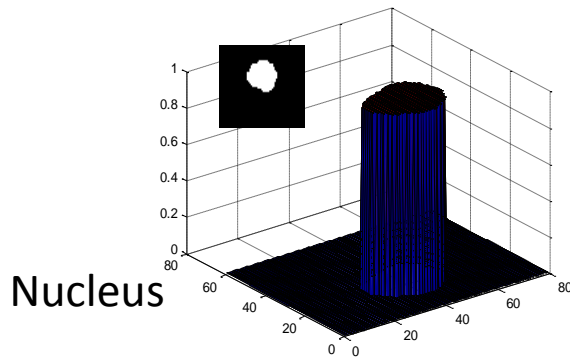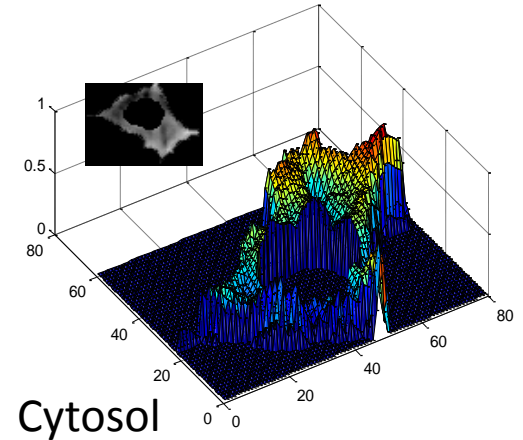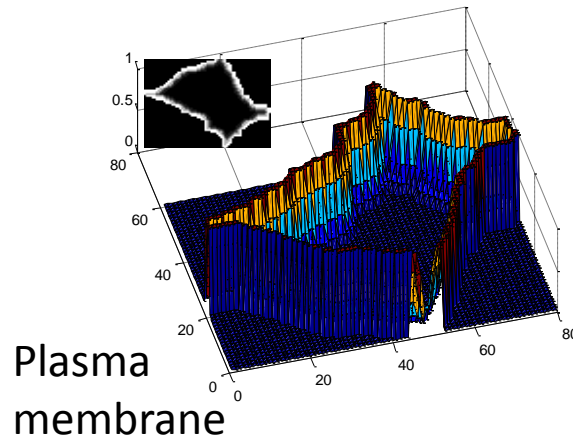Raw image   Cell region   Actin   GFP   Nucleus

Actin

Plasma membrane

Cytosol

Nucleus

Cytoplasm
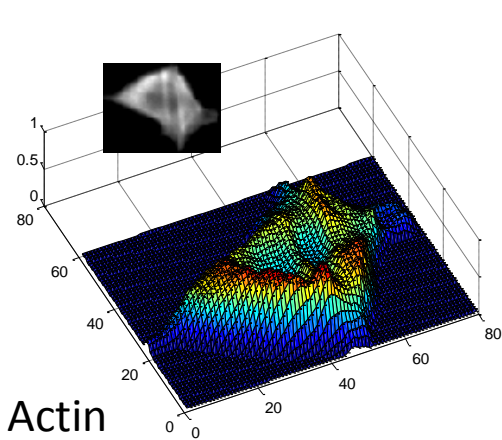
Transfection and sample preparation

- **HeLa cell line**

- **96 Invitrogen GFP-tagged cytoskeletal protein constructs**

- **Transfection reagent: Lipofectamine2000**

- **Concentration: 10ng/ul**

- **Dyes:**
  - Nucleus: Hoechst 33342
  - Actin: Texas Red Phalloidin

Transfection and sample preparation

|	1	2	3	4	5	6	7	8	9	10	11	12
A	DCAMKL1	CORO2B	PDLIM3	WASPIP	KARCA1	CTTN	VIL1	ACTB	WASL	DNM2	ZYX	TUBA6
B	CDC42	CFL1	TGOLN2	TAGLN	WAS	EVL	TUBE1	TNS	CLIPR-59	DSTN	ITGB2	CORO1B
C	ATP6V1C1	TAGLN3	VIL2	TUBD1	ATP5G2	TUBGCP3	PARVG	RDX	ITGB3BP	FSCN3	TEKT3	PAK4
D	GJB2	KALPHA1	CAPZA3	ACTG1	KPTN	TEKT1	PTK9	KIF3A	TPM1	TUBB2	ITGB1	ADAM15
E	CAPN1	ACTA2	CETN2	KRT18	CAPG	PLS1	VASP	HTATIP	ARHGEF6	WASF3	KIAA0555	EIF2C1
F	LPXN	ATP1B3	VAMP4	ACTB	PXN	DCTN1	WASF2	PXN	MSN	KIF2C	ITGB7	CAV3
G	PLS3	ADRM1	ARPC1B	TUBG1	MRLC2	GTSE1	TAGLN2	ARPC5	ACTN2	MYO3A	ACTN1	FILIP1
H	ARP3BETA	CNN3	LCP1	VIM	NINJ2	PFN2	PARVA	TPM2	MYO1A	ANLN	FSCN1	KRT8

# Transfection and sample preparation
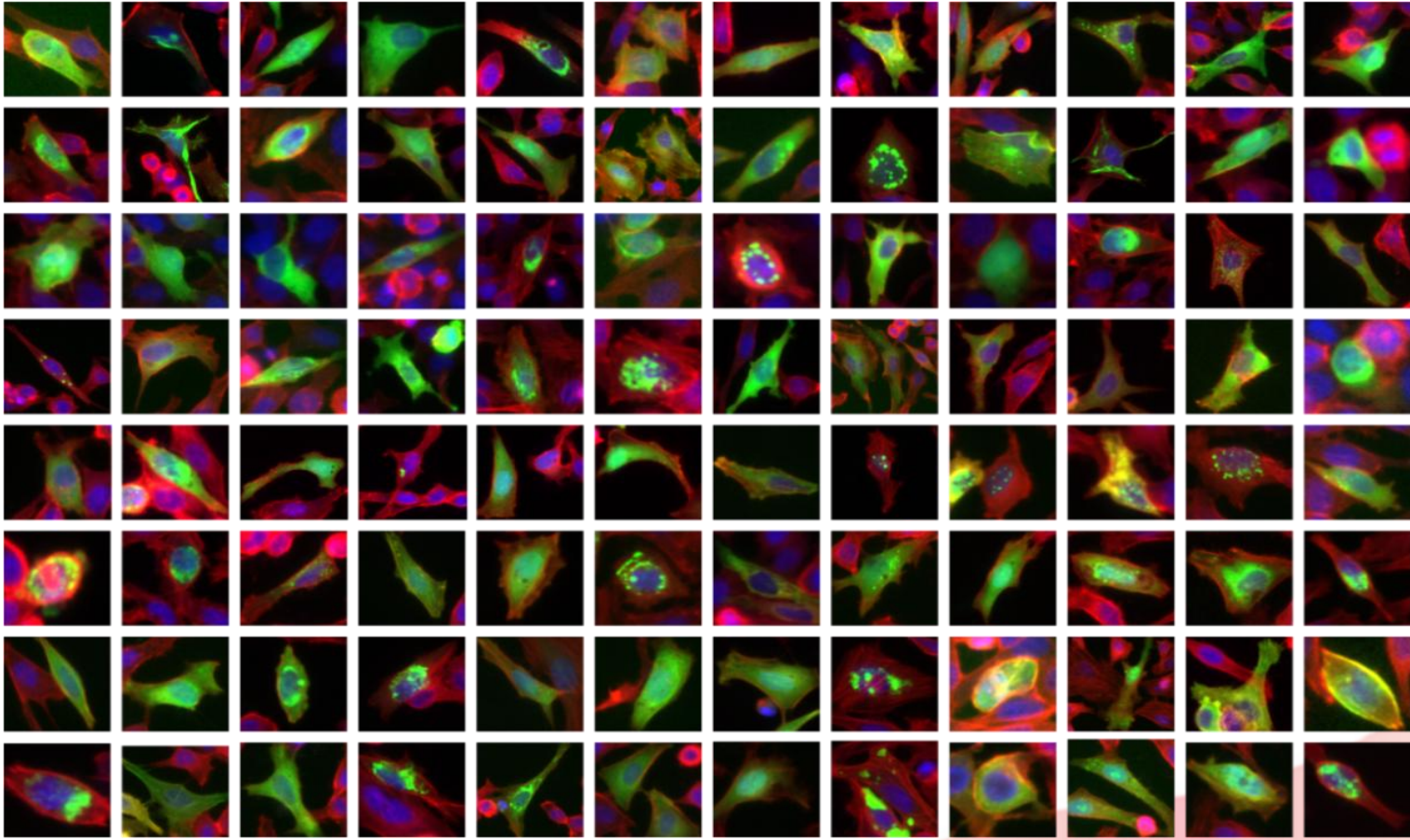
2D Image Acquisiton

- **Cellomics vHCS: Scan V Target Activation application system**
  - 20X magnification

- **Images of 7 constructs are lost due to system problems**
- **The image set contains information of 89 constructs**
  - ~7000 images
  - Three channels:
    - Blue-Nuclei
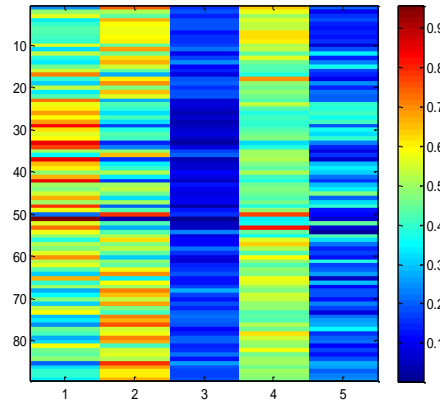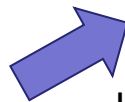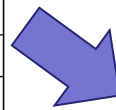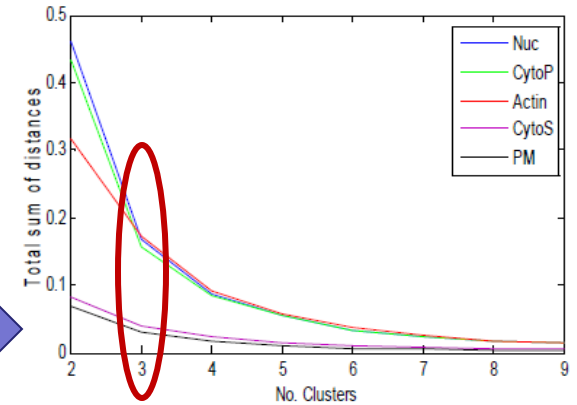   - Red-Actin
    - Green-Cytoskeletal protein
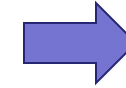
Statistic Analysis
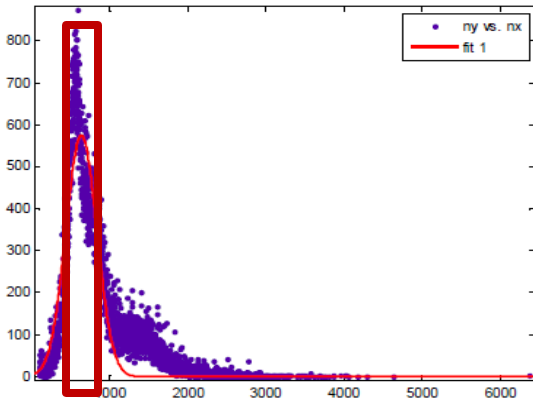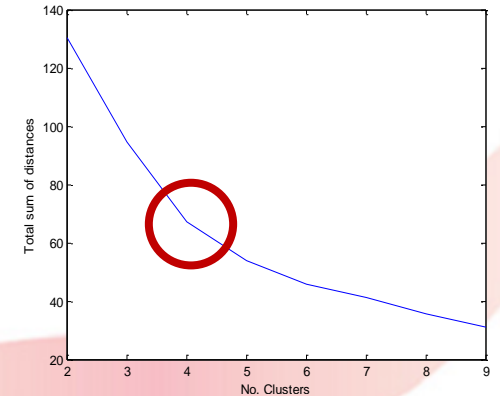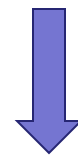

Cell selection
based on nuclear size
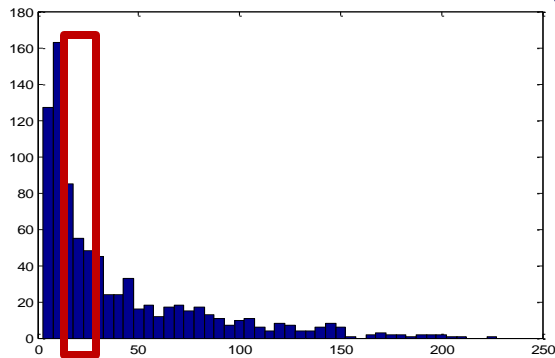
Cell selection
based on GFP intensity

Heat map of colocalization value:
89 proteins X 5 subcellular compartment

Colocalization indexing
with k-means clustering

Protein clustering
based on colocalization indexing
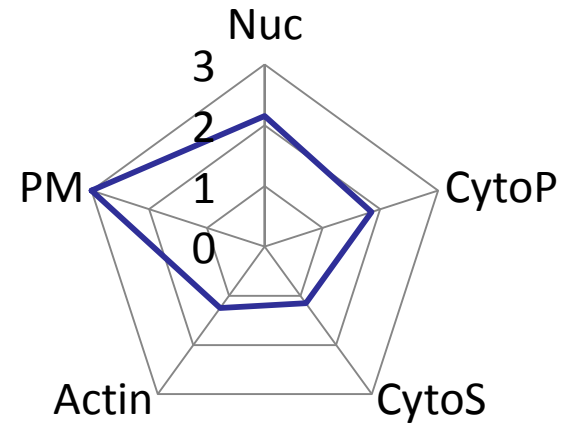
Colocalization of Protein Clusters

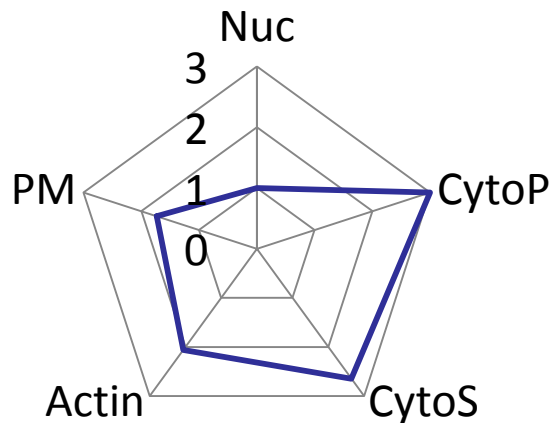|	*Nucleus*	*CytoP*	*CytoS*	*Actin*	*PM*	*protein No.*
Cluster1	52.42% ±4.54%	47.58% ±4.54%	16.26% ±3.31%	49.40% ±5.41%	14.37% ±5.81%	21
Cluster2	56.63% ±8.48%	43.37% ±8.48%	7.85% ±2.66%	43.96% ±4.66%	**37.60% ±4.18%**	21
Cluster3	34.70% ±6.08%	**65.30% ±6.08%**	**18.62% ±3.36%**	**53.23% ±7.53%**	20.54% ±5.50%	34
Cluster4	**75.72% ±9.07%**	24.28% ±9.07%	6.81% ±3.07%	45.99% ±13.04%	15.57% ±7.27%	13

Colocalization of Protein Clusters
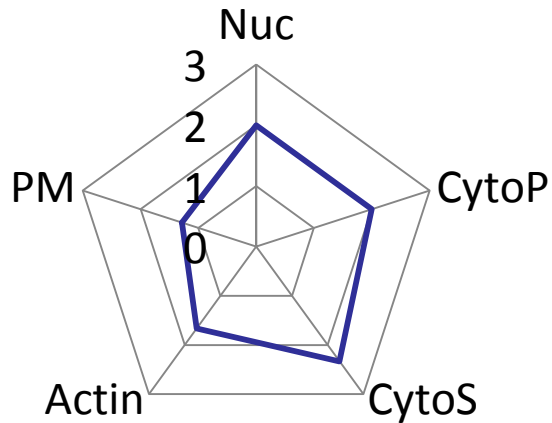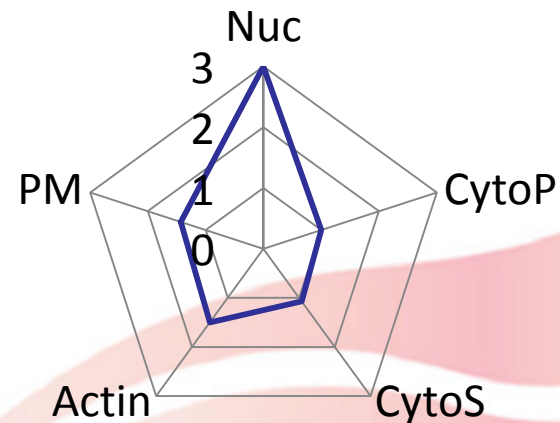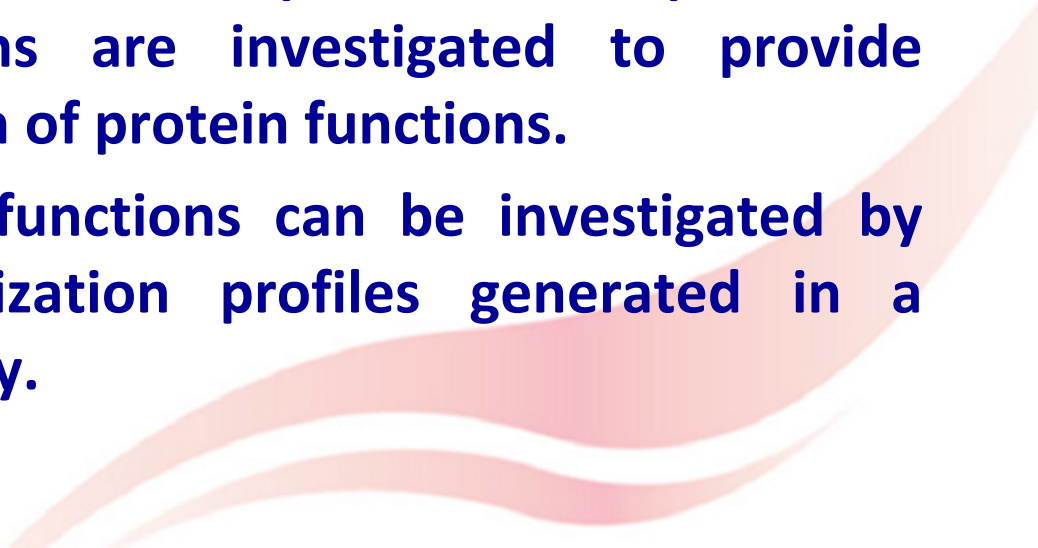
Conclusion

- We developed a computational framework and optimized every step in the framework to quantify the subcellular localization of cytoskeletal proteins with a single colocalization measurement.

- The framework is applied on a two-dimensional image set containing images of 89 cytoskeletal protein. The subcellular localizations of those cytoskeletal proteins are quantified and localization patterns are investigated to provide references in investigation of protein functions.

- Proteins with unknown functions can be investigated by comparing with colocalization profiles generated in a cytoskeletal protein library.

Acknowledgement

- **All the members in Rajapakse Lab**

- **All the members in Matsudaira Lab**

- **This work was partly supported by Computation and Systems Biology programme of the Singapore-MIT Alliance.**