Quantifying Caveolin-1 Molecular Domains via Machine Learning and Computational Network Analysis of Super-resolution Microscopy Data

Ismail M. Khater¹, Fanrui Meng², Ivan Robert Nabi², Ghassan Hamarneh¹

¹Medical Image Analysis Lab, Simon Fraser University
²Life Sciences Institute, University of British Columbia

Bringing the ASCB to the LSI
December 8th 2016
Quantifying Caveolin-1 Molecular Domains via Machine Learning and Computational Network Analysis of Super-resolution Microscopy Data
Quantifying Caveolin-1 Molecular Domains via Machine Learning and Computational Network Analysis of Super-resolution Microscopy Data
Motivation

● Design a computational pipeline for detecting and identifying subcellular structures from super-resolution data automatically

Challenges

● Big data: hundreds of thousands or millions of blinks
● Noise: false positives & localization error
● Limited work on automatic & quantitative 3D methods
Motivation

- Design a computational pipeline for **detecting and identifying** subcellular structures from super-resolution data automatically.

Challenges

- **Big data**: hundreds of thousands or millions of blinks
- **Noise**: false positives & localization error
- **Limited work** on automatic & quantitative 3D methods
Overview

● Structures of interest (caveolae and scaffolds)
● Proposed Method
  Preprocessing | Filtering | Segmentation | Analysis | Identification
● Our Data
● Results & Conclusions
Structures of Interest

Caveolin-1 is overexpressed and secreted in prostate tumors

Caveolae: Ω-shaped (50-100 nm) invaginations

Scaffolds: Non-Caveolar domains
SMLM’s super-resolution enables studying structures

Caveolin-1 in PC3 Cell lines

Widefield (TIRF)
SMLM’s super-resolution enables studying structures

Caveolin-1 in PC3 Cell lines

Widefield (TIRF) GSD (TIRF)
Proposed Method: cluster extraction and description

Analyses super-resolution data using the following steps:

1. GSD image
2. 3D point cloud
3. ROIs

which clusters are caveolar and which are scaffolds?

Identification

Feature Extraction

Network Statistical

Hollowness Shape

Segmentation Filtering

Network analysis
Proposed Method: cluster identification

Dataset
PC3 cells
PC3-PTRF cells

Training phase

Filtering and blobs generation
Learn groups
Match groups / identification
Labeling

Test image
Filtering & blobs generation
Find the most similar group to each one of the blobs
Blobs identification
Proposed Method - features

● Fast
  Processes ~1 million blinks in ~1 minute (on 3.4 GHz 32GB RAM)

● Scalable
  Parallelizable and runs on a cluster of compute nodes

● 3D from Start to Finish
  All analysis is in 3D: point cloud, network model, segmentations, detection, features

● Robust to Noise
  Avoids ‘fake’ clusters, those with features similar random networks

● Flexible & General
  Extracts heterogeneous clusters with different shapes and sizes

● Quantitative & Discriminatory
  Identifies different types of clusters via quantitative molecule- and cluster-level features
## Dataset

- 4 experiments of PC3 cell lines

<table>
<thead>
<tr>
<th>Exp. #</th>
<th>PC3 cells</th>
<th>PC3-PTRF cells</th>
<th>PTRF mask</th>
<th>Used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. 1</td>
<td>11</td>
<td>11</td>
<td>No</td>
<td>Training</td>
</tr>
<tr>
<td>Exp. 2</td>
<td>10</td>
<td>10</td>
<td>No</td>
<td>Training</td>
</tr>
<tr>
<td>Exp. 3</td>
<td>11</td>
<td>10</td>
<td>No</td>
<td>Training</td>
</tr>
<tr>
<td>Exp. 4</td>
<td>9</td>
<td>10</td>
<td>Yes</td>
<td>Testing</td>
</tr>
</tbody>
</table>
Results - Clusters

Blobs identification and labeling in both populations

PC3

Unlabeled

Labeled / identified blobs

PC3-PTRF

Scaffolds

Caveolae
Results - Caveolae (GB)

A) The blob's blinks

B) Blob's network @ thresh. = 80 nm
   # connected components = 1

C) Blob's network @ thresh. = 80 nm
   # modules = 6

D) Surface reconstruction

E) Cross sections
Results - Scaffold S2 (GA)

A) The blob's blinks

B) Blob's network @ thresh. = 50 nm
   # connected components = 1

C) Blob's network @ thresh. = 80 nm
   # modules = 3

D) Surface reconstruction

E) Cross sections

- blinks
- centroid
Results - Hollowness and Sphericity

A) Avg. distance to centroid (nm)

B) CS

- PC3-PTRF groups
- PC3 groups
Results - Hollowness and Sphericity

A) Caveolae (hollow structures) & Scaffolds (less spherical structures)

G1 is similar to GC & GD in hollowness

B) PC3-PTRF groups & PC3 groups
Results - Modularity Analysis

A) # conn. comp. for each blob at various thresholds

B) # modules for each blob at various thresholds
Results - Modularity Analysis

A) 

All the blinks got connected after 50 nm

B) 

Plateau, # modules get fixed after 50 nm
Conclusion

We presented:

● Fast and scalable 3D clustering method for analyzing super-resolution data
● Machine learning and network analysis to identify biological structures
● A system that provides quantification at the molecular level as well as at the cluster level

Future work:
Share our tools and test on other biological structures
Thank You!

contact: ikhater@sfu.ca
Backup slides
### Related Work

<table>
<thead>
<tr>
<th>Work</th>
<th>Technique</th>
<th>Weakness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Owen et al. 2010</td>
<td>Ripley’s K, L, H-functions for cluster analysis</td>
<td>Global cluster analysis</td>
</tr>
<tr>
<td>Pageon et al. 2013</td>
<td></td>
<td>Restricted to analyze homogeneous clusters</td>
</tr>
<tr>
<td>Rossy et al. 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pereira et al. 2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lillemeier et al. 2010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lagache et al. 2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>El Beheiry et al. 2013</td>
<td>Visualization, reconstruction, and density plot of 3D super-resolution data</td>
<td>Limited in quantification capabilities</td>
</tr>
<tr>
<td>(ViSP)</td>
<td></td>
<td>Assumes the data is clean</td>
</tr>
<tr>
<td>Rossy et al. 2014</td>
<td>Univariate/bivariate Getis and Franklin’s local point pattern analysis</td>
<td>Model-free method (difficult to judge its performance and its results)</td>
</tr>
</tbody>
</table>
## Related Work continued

<table>
<thead>
<tr>
<th>Work</th>
<th>Technique</th>
<th>Weakness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caetano et al. 2015 (MIiSR)</td>
<td>Density-based methods and Ripley’s H-function</td>
<td>Cannot deal with varying clusters densities and hollow clusters</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitive to noisy events</td>
</tr>
<tr>
<td>Rubin-Delanchy et al. 2015</td>
<td>Bayesian approach and Ripley’s functions</td>
<td>Sensitive to the prior settings</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not suitable for small clusters</td>
</tr>
<tr>
<td>Levet et al. 2015 (SR-Tesseler)</td>
<td>Voronoi tessellation</td>
<td>Crude segmentation</td>
</tr>
<tr>
<td>Andronov et al. 2016 (ClusterViSu)</td>
<td></td>
<td>Hard to deal with hollow clusters</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It’s multiscale capability are limited</td>
</tr>
</tbody>
</table>
Proposed Method

Pipeline:
Preprocessing
Filtering
Segmentation
Analysis
Identification

GSD image → 3D point cloud → ROIs → Network analysis

Identification
Feature Extraction
Statistical
Network
Hollowness
Shape

Segmentation
Filtering
Results - Merging Algorithm

Get rid of the multiple blinking of individual fluorophores (artifact) via Iterative merging algorithm

Before merge

After merge

PC3-PTRF

degree

Unweighted degree

# blinks (%)
Results - Clusters Extraction

A) 3D point cloud

B) Multi-threshold network analysis

C) Filtering

D) Blobs segmentation
Results - Hollowness and Sphericity

A) Caveolae (hollow structures)

G1 is similar to GC & GD in hollowness

B) Scaffolds (less spherical structures)

Bar charts showing the average distance to centroid and FA for different groups (GA, G2, GB, GC, GD, G1).
Results - Size Measures

- Volume (nm$^3$)
- X range (nm)
- Y range (nm)
- Z range (nm)

Comparing PC3-PTRF and PC3 groups.
Results - Network Measures

[Graphs showing network measures for different groups, including number of blinks, average degree, average characteristic path length, and modularity.]
Results - Modularity Analysis

Avg. modules sizes for each blob at various thresholds:

- PC3-PTRF GA
- PC3-PTRF GB
- PC3-PTRF GC
- PC3-PTRF GD

Avg. conn. comp. sizes for each blob at various thresholds:
Key Features

Our Method Key Features:

- Very fast
- Extract heterogeneous clusters
- Scalable and can be applied to the whole cell (Not just an ROI)
- Avoid fake clusters
- 2D/3D cluster analysis
- Quantification at the molecular level
Method Key Features - Fake Clusters

3D point cloud

Other clustering methods

Our clustering method

34
Proposed Method

- Very fast
- Scalable
- Avoids ‘fake' clusters
- 3D cluster analysis
- Extract heterogeneous clusters

![Diagram showing the proposed method steps: GSD image, 3D point cloud, ROIs, Network analysis, Segmentation, Filtering]
Proposed Method

Quantification at the molecular level & identification

GSD image → 3D point cloud → ROIs → Network analysis

Identification

Hollowness
Network
Statistical

Shape
Feature Extraction
Segmentation
Filtering