

KNIME @ HiTIF: Bioimaging Workflows for Looking Inside Cells

Prabhakar Reddy Gudla
High-Throughput Imaging Facility (HiTIF)

KNIME Fall Summit, Austin, TX
Nov 1 – 3, 2017

Overview

- High-throughput Imaging (HTI)
- Why KNIME?
- KNIME (KNIP) HTI Applications
 - Transcription dynamics from 2D-t images
 - Deep learning for object detection

HTI Enables Systematic Study of Cell Function

Investigator

Experimental treatment

Imaging-based cellular assay

Phenotypic change

HiTIF

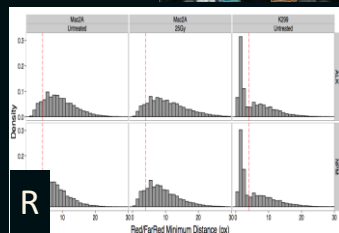
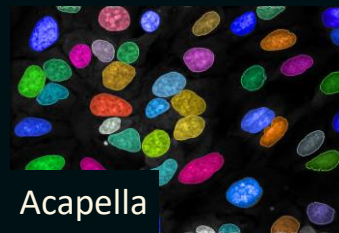
Automated liquid handling



High-throughput microscopy



High-content image analysis



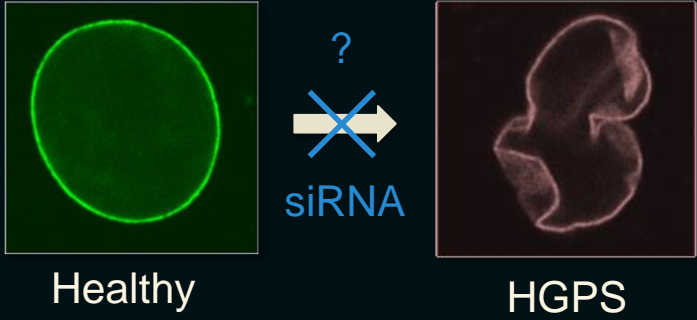
Up to:

- 10^4 Wells
- 10^4 Cells/Well
- 10^2 Features/Cell

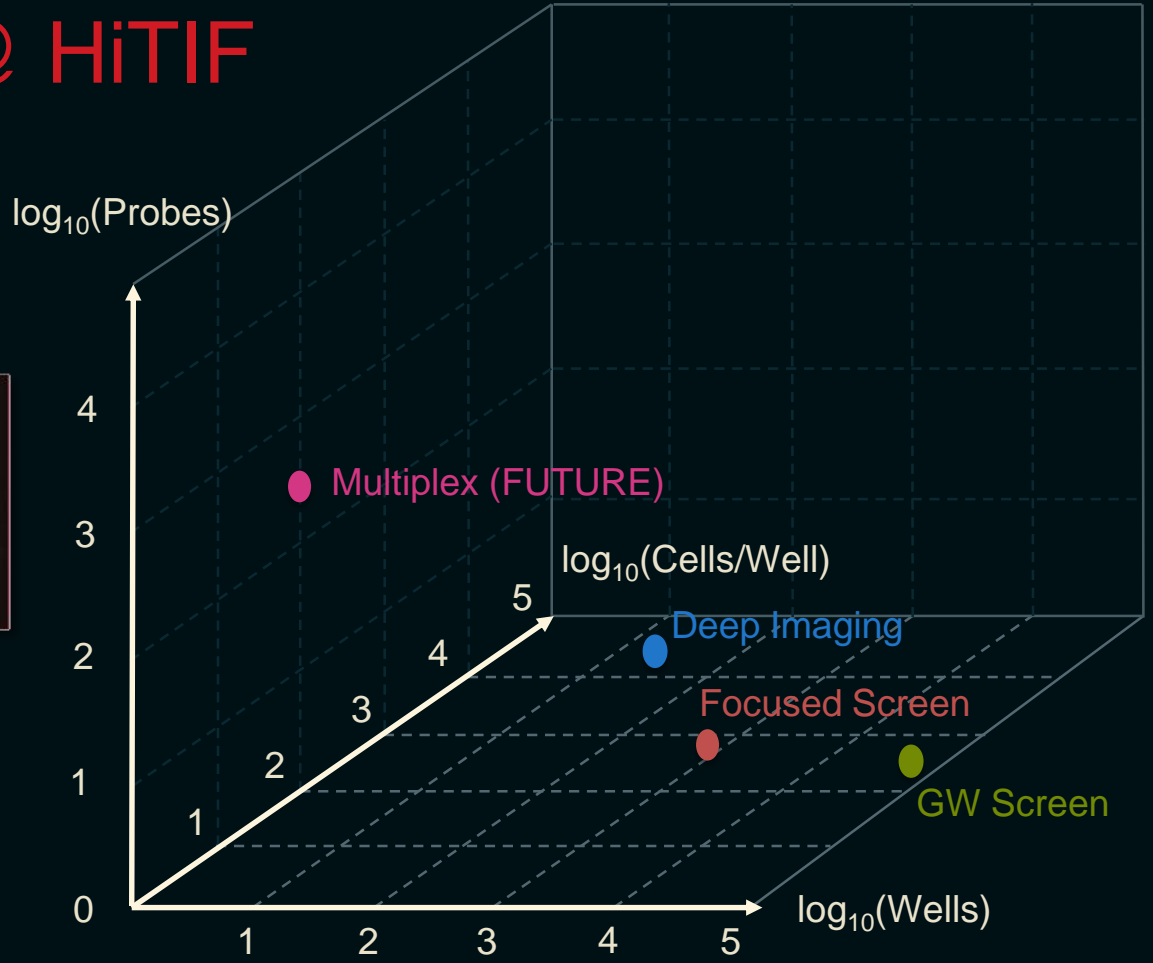
50 GB – 2 TB

Assays Formats @ HiTIF

Focused Screen



Deep Imaging



Why KNIME?

Reproducible

Open source

Scalable



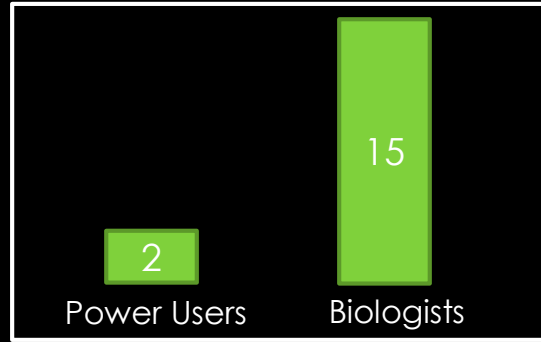
Source: Jan Eglinger, KNIME Summit, Spring 2017



Computational Infrastructure

OLD Stack

Columbus + Acapella
+
R/Matlab
+
ImageJ/Fiji



NEW Stack

KNIME + KNIME Server
+
R/Python/Java

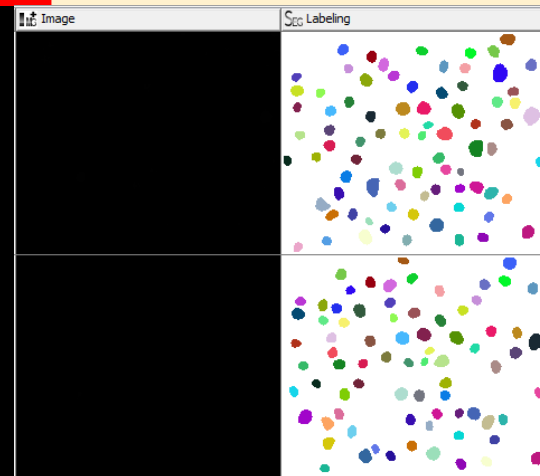
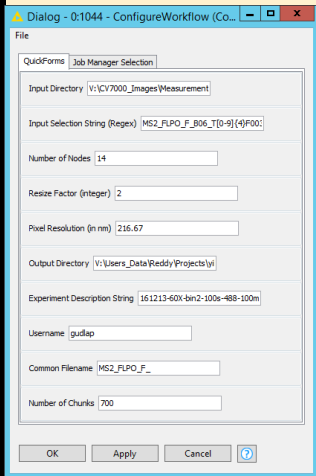
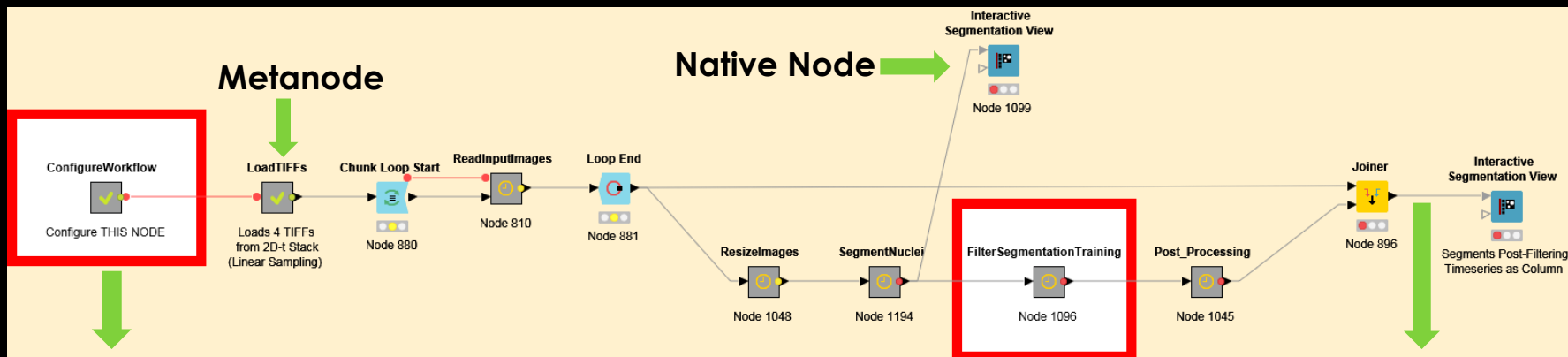
• Hardware

- 2 X 16-Core AMD, 256 GB RAM, 1.5 TB SAN, Windows Server 2012
- HPC Batch Cluster: Intel 28 Core (w/ HT), 256 GB, 4 K80 GPU, 800 GB SSD
 - Batch Limit: 3072 CPUs, 32 GPUs, and 10 days
 - Singularity (Container technology for HPC)

• Storage

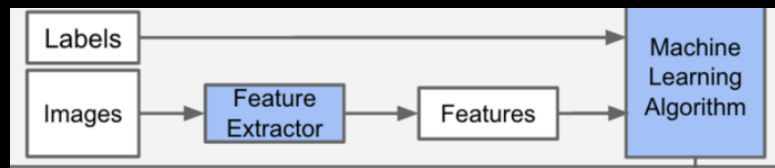
- 72 TB Tier2 Isilon (Perkin Elmer's Columbus/OMERO)
- 60 TB Tier3 for archiving

KNIME @ HiTIF: Design Philosophy



KNIME @ HiTIF: Data Driven

Interactive



30-60 mins

Batch Mode

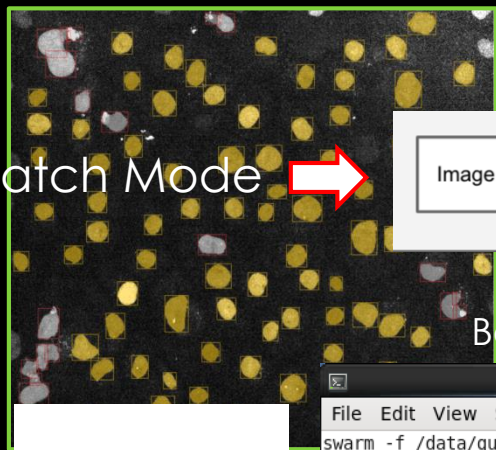


Features

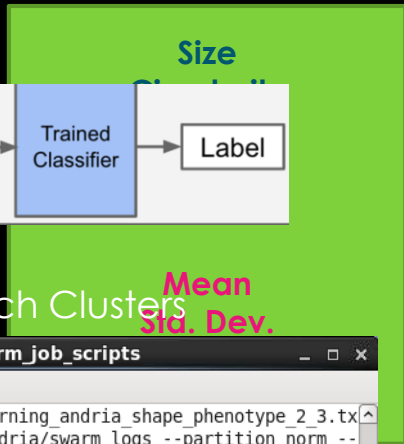


Batch for HPC Batch Clusters

```
gudlap@biowulf:/data/gudlap/andria/swarm_job_scripts
File Edit View Search Terminal Help
swarm -f /data/gudlap/andria/swarm_job_scripts/knime_rflerning_andria_shape_phenotype_2_3.tx
t -g 64 -t 56 --gres=lscratch:300 --logdir /data/gudlap/andria/swarm_logs --partition norm --
time 06:00:00 --sbatch "--exclusive"
### Swarm, 64G, 56 threads, ccr partition, exclusive node mode, 768 flex files per parallel f
ork, 56 parallel forks, 56 MaxThreads, walltime 6 hours
## Plates: HTIF00023_2
swarm -f /data/gudlap/andria/swarm_job_scripts/knime_rflerning_andria_shape_phenotype_2_2.tx
t -g 64 -t 56 --gres=lscratch:300 --logdir /data/gudlap/andria/swarm_logs --partition ccr --t
ime 06:00:00 --sbatch "--exclusive"
```



File	Hitte
V:\CV\7000_Imag...	Good
V:\CV\7000_Imag...	Good
V:\CV\7000_Imag...	Good
V:\CV\7000_Imag...	Good
V:\CV\7000_Imag...	Good
V:\CV\7000_Imag...	Good
V:\CV\7000_Imag...	Good
V:\CV\7000_Imag...	Good
V:\CV\7000_Imag...	Good
V:\CV\7000_Imag...	Good



ConfigureWorkflow

Configure THIS NODE

Write V...

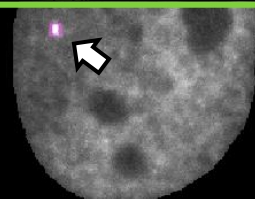
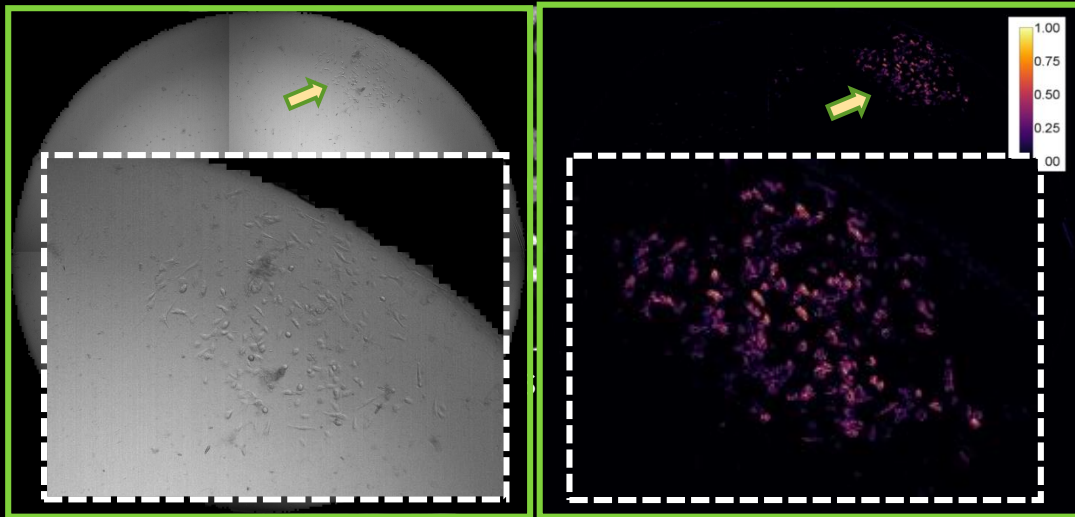
SAVE

Table Writer

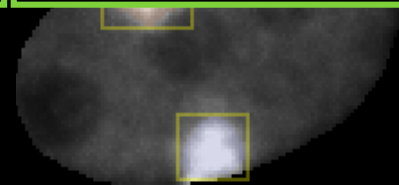
USE

KNIME @ HiTIF

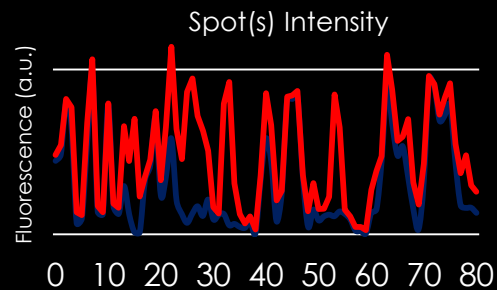
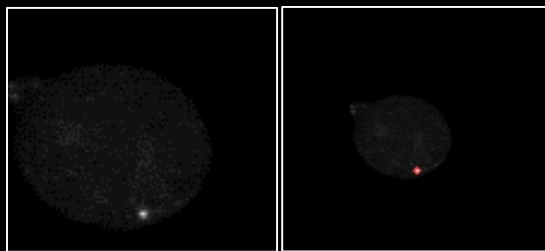
- Segmentation
 - Bright-field cell segmentation
 - Nucleus Segmentation
 - Chromosome Territories
 - FISH Spots/Transcription Sites
- Tracking (2D-t)
 - Live cell imaging
 - Tracking transcription site(s)
- Registration (Affine)
- Phenotype/Outlier Detection



FISH



Chromosome Territories

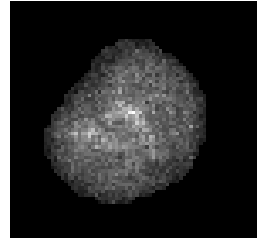
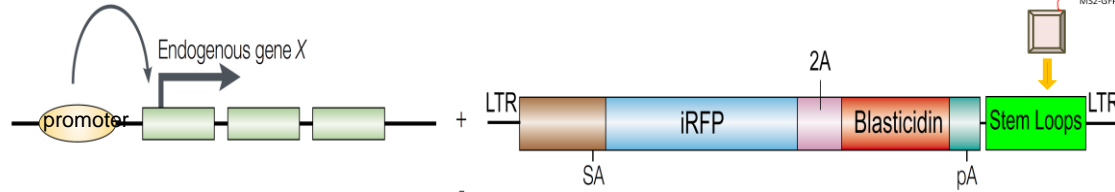
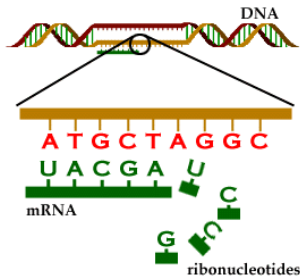


Overview

- High-throughput Imaging (HTI)
- Why KNIME?
- KNIME Applications
 - Transcription dynamics from 2D-t images
 - Deep learning for object detection

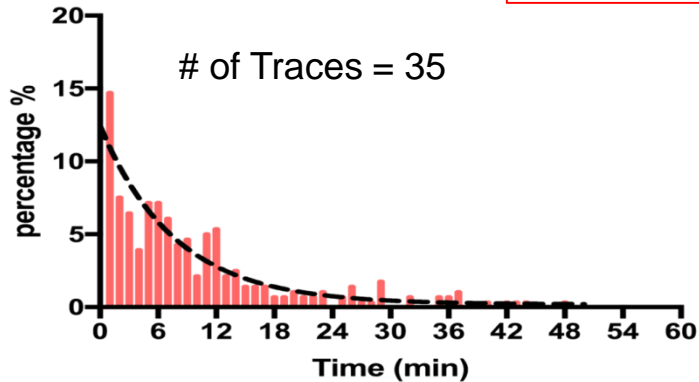
Gene-Trap: Transcription Dynamics of Thousands of Genes

Transcription



Frame Interval: 100 s
Expt. Time: 16 hrs
(HBE Cell)

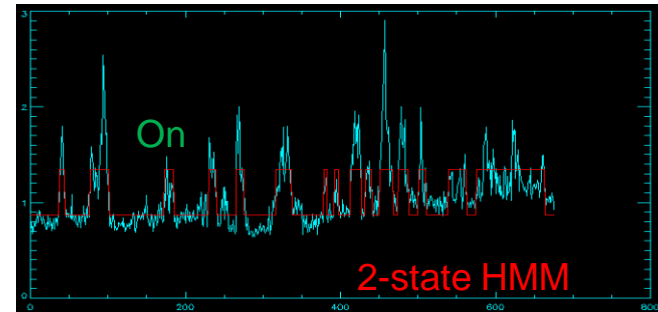
1 gene, MANUAL PROCESS, 6 mo



Exponential fit to On-time Histogram



Fluorescence (a.u.)



0 hr

Time

5 hr

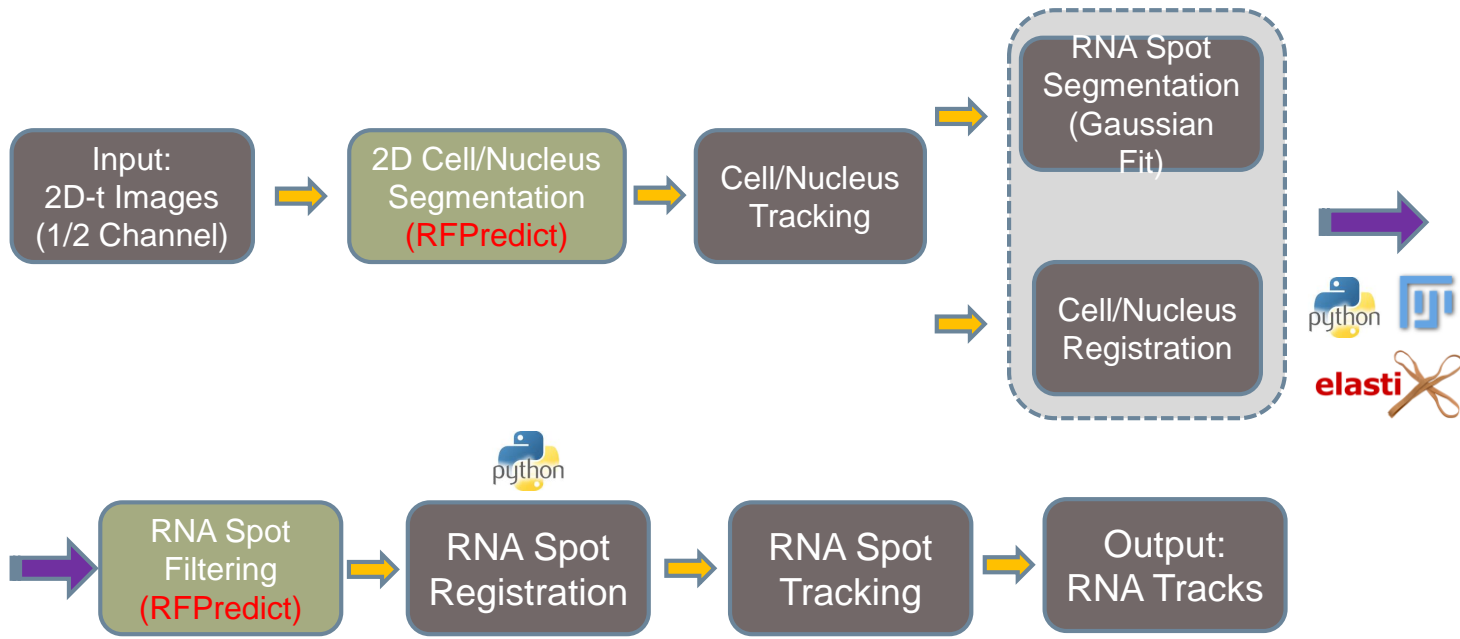
Image(s): Yihan Wan, Larson

KNIME Automated Workflow

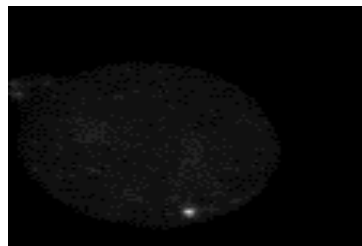
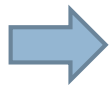
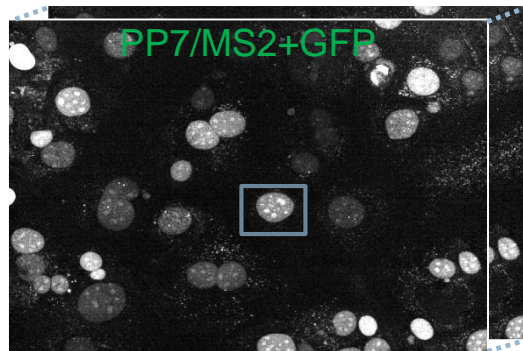
Yokogawa CV7000S



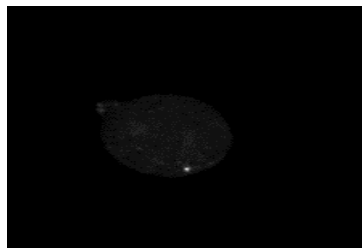
6 Wells, 6 Fields
Frame Interval: 100 s
~ 20, 000 images
1276x1076 pixels



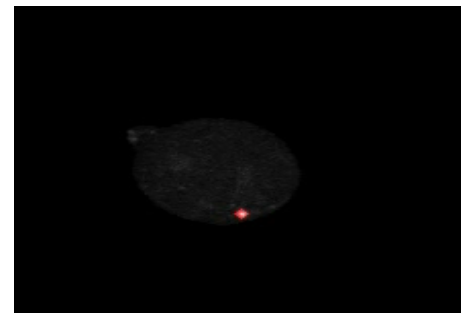
Tracking Multiple Transcription Sites



Unregistered Ch0
PP7/MS2+GFP

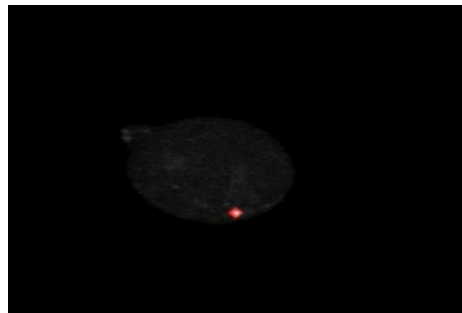


Registered Ch0



Ch0 with Tracks
(Registered)

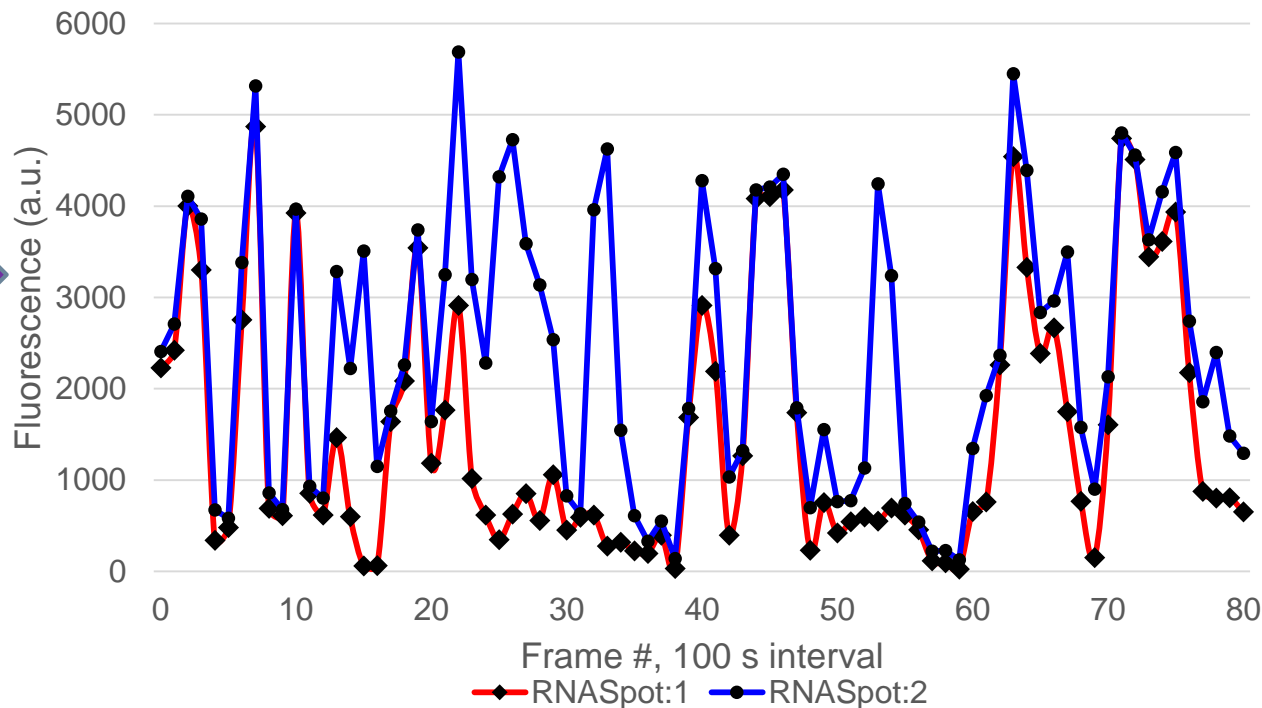
Tracking Multiple Transcription Sites



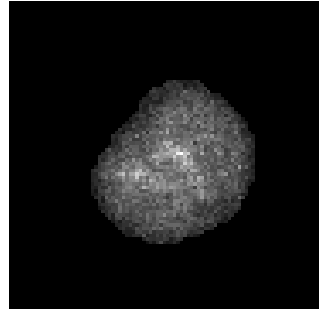
Ch0 with Tracks
(Registered)



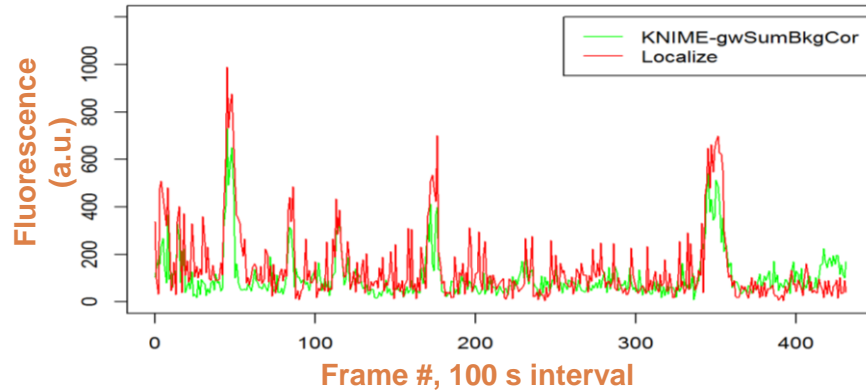
Spot(s) Intensity from Unregistered Images



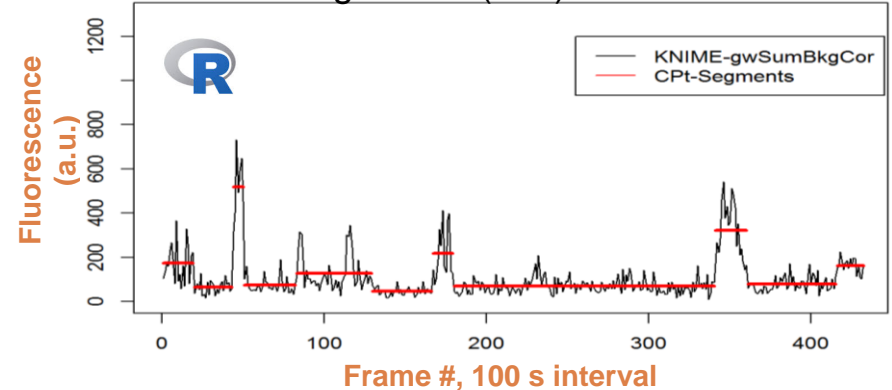
Modeling Transcription Bursts



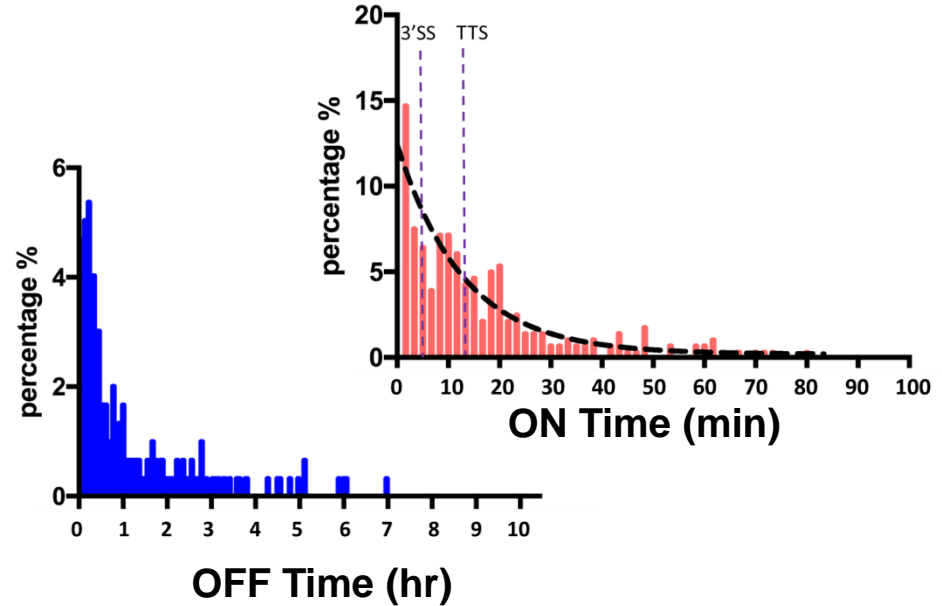
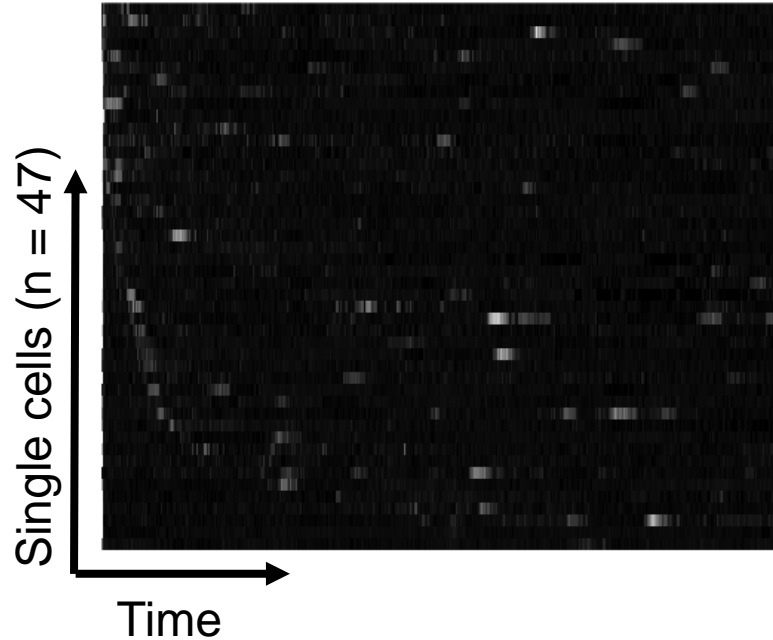
Automated vs Manual



Change Point (Cpt) Detection/HMM



Visualizing Transcription and Splicing Dynamics of ERRFI1



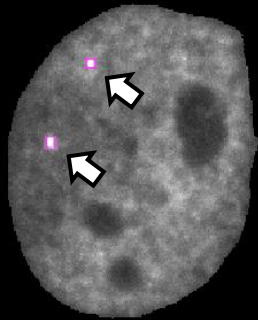
Status → 63 Clones → 23 Genes → 3-4 mo

Overview

- High-throughput Imaging (HTI)
- Why KNIME?
- KNIME Applications
 - Transcription dynamics from 2D-t images
 - Deep learning for object detection

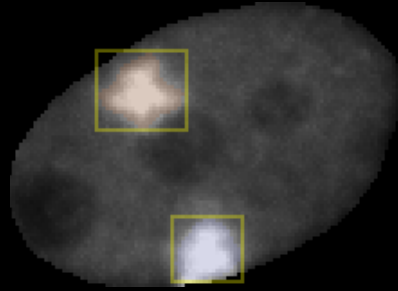
Deep learning for detecting subcellular structures

~10-25 nm
(3 x 3 pixels)*



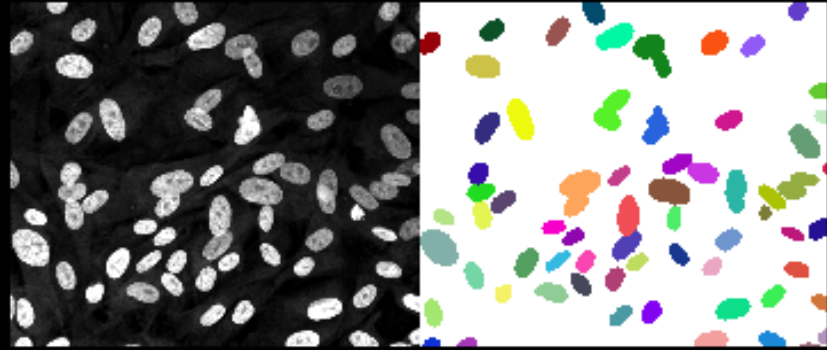
DNA FISH

~2-5 μm
(32 x 32 pixels)*



Chromosome Territories

~15-30 μm
(100 x 100 pixels)*



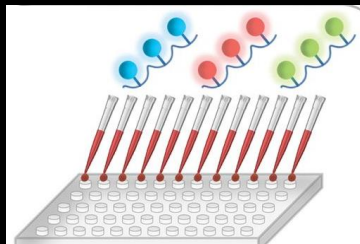
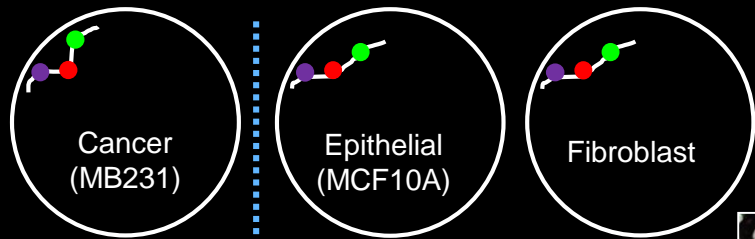
Nucleus

Use KNIME+KNIP for generating data for DL networks

Speed, Accuracy, and “NO PARAMETERS TWEAKING”

* 40X Dry Objective

High-throughput Position Mapping (HIPMap)

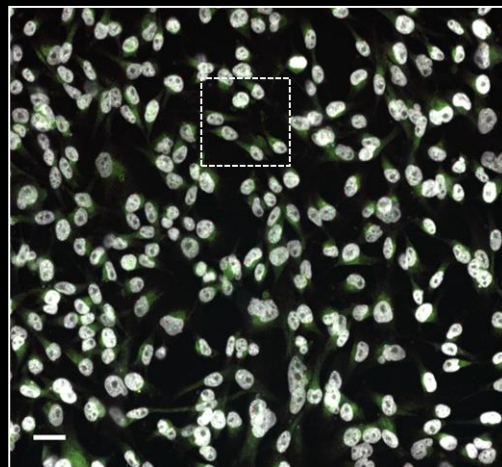


Oligo-paint in 384-well plate

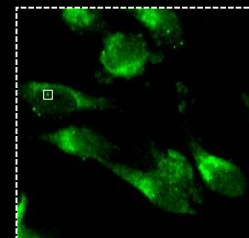
Genomic Loci: 189

Study: 50 plates ($\sim 10^5$ cells/plate)

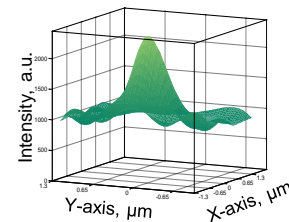
Conditions: Normoxia vs. Hypoxia



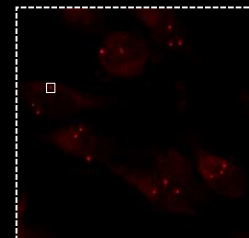
DNA FISH using Oligo-Paint (MB231)



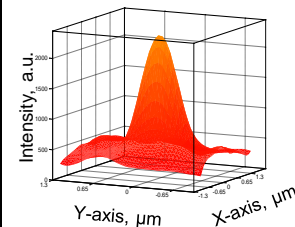
Alexa488 Channel



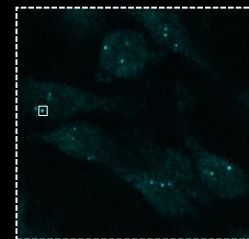
SNR ~ 2



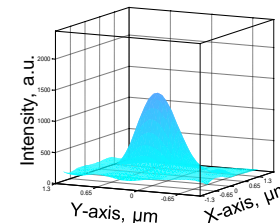
ATTO565 Channel



SNR ~ 5



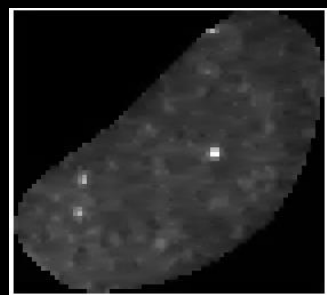
Cy5 Channel



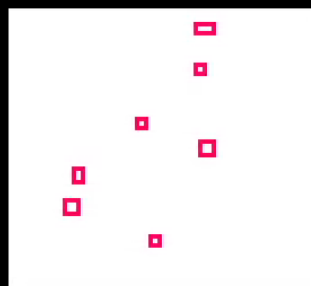
SNR ~ 9

DNA FISH Spot Detection

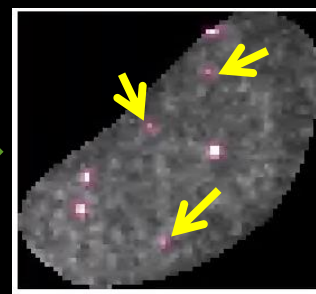
Crop FISH Channels ← Segment Nuclei (DAPI) ← MIP



Greyscale FISH Image
(Far Red)



Optimized (Weak)
Spot Detection



Overlay
(False Positives)



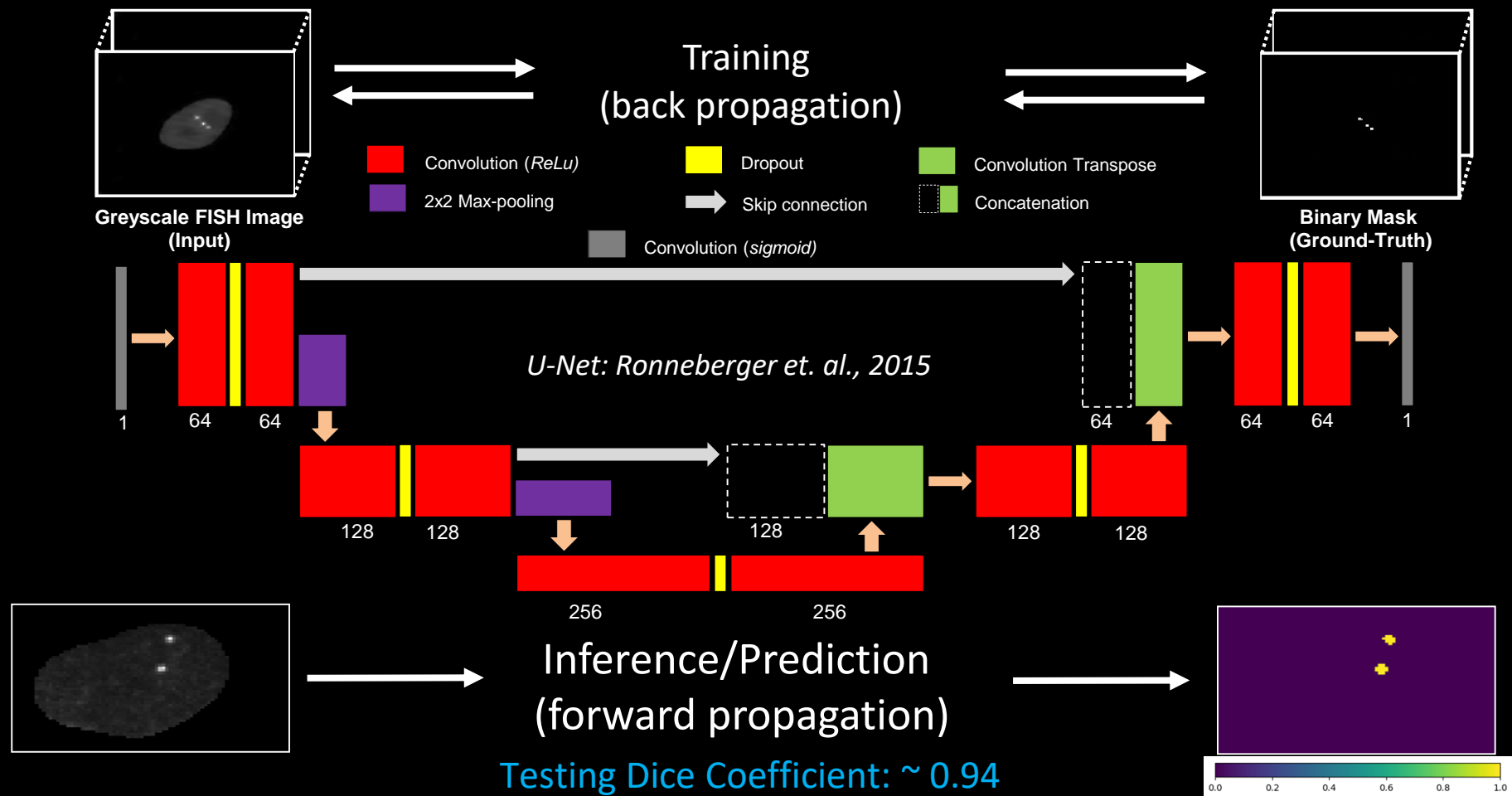
Machine
Learning
(ML)

FISH Image (SNR > 5.0)



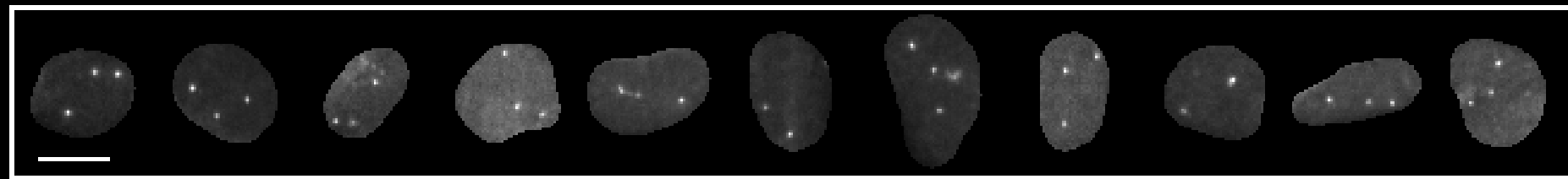
Poor features spot-like objects → Required separate ML model/channel

U-Net_{2L}: DNA FISH Spot Segmentation

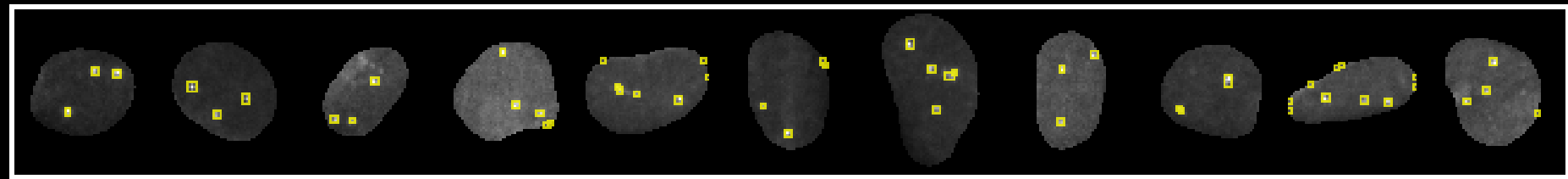


KNIME for Generating UNet_{2L} Training Data

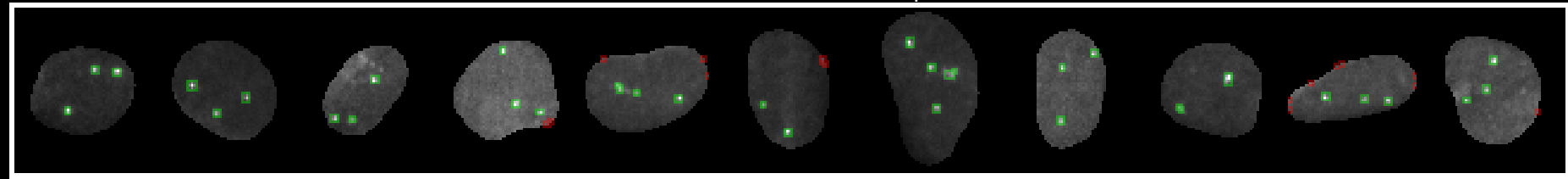
FISH Images From Three Spectral Channels (Plate Optimization Step)



Detect Spots Using Wavelets

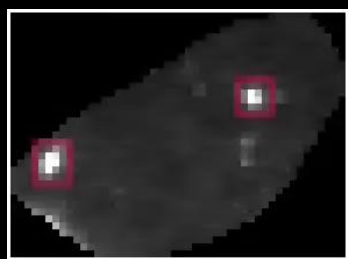
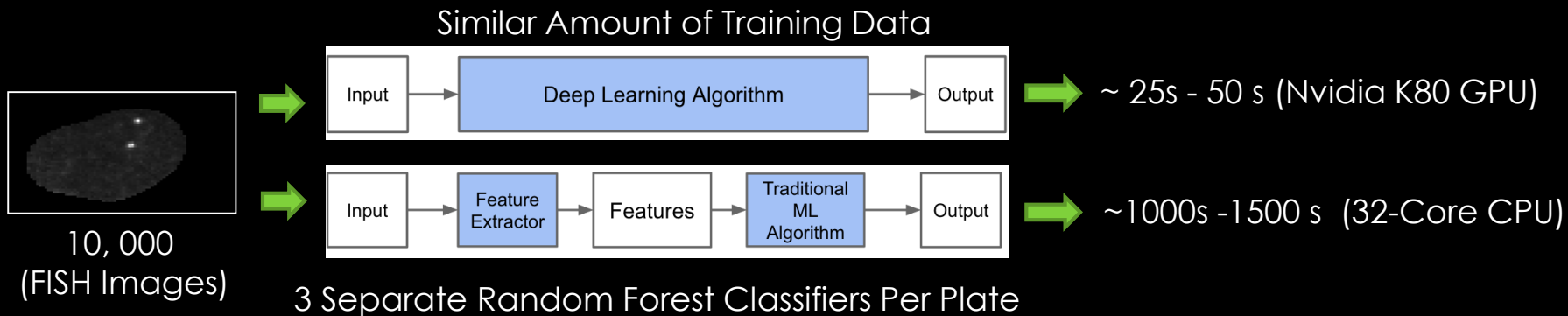


Annotate "Good" Spots

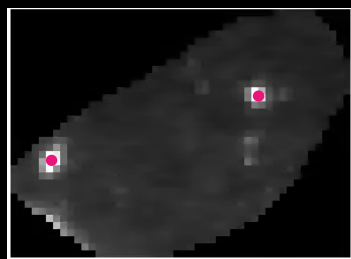


Training Set: 189 FISH Images Validation Set: 23 FISH Images

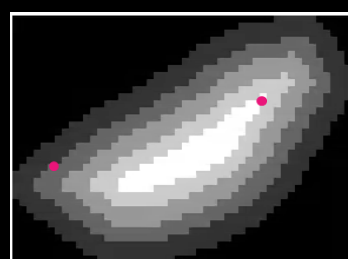
UNet_{2L} Performance



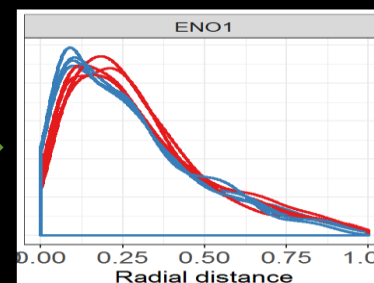
Threshold + Label
(4-Connected)



Centroid (Binary Mask)
CoG (Mask+Intensity)

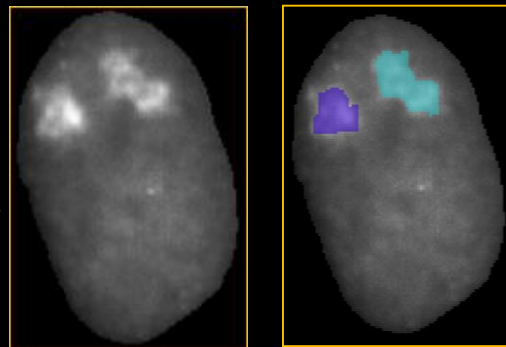
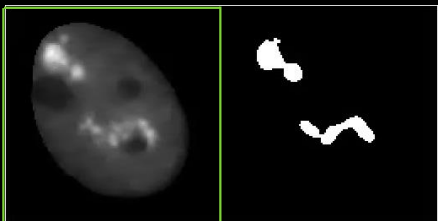
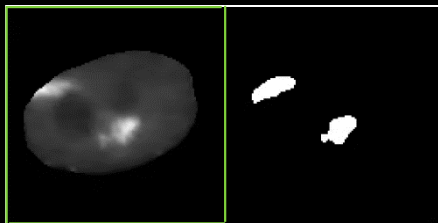
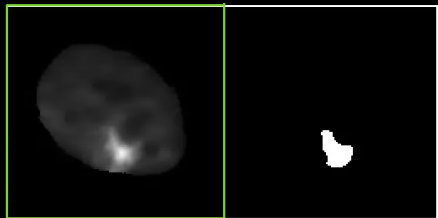


Equidistant/Area Shells



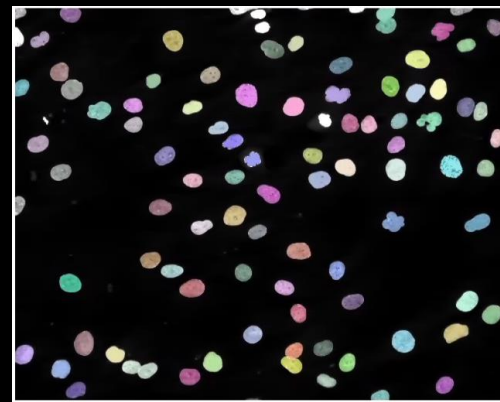
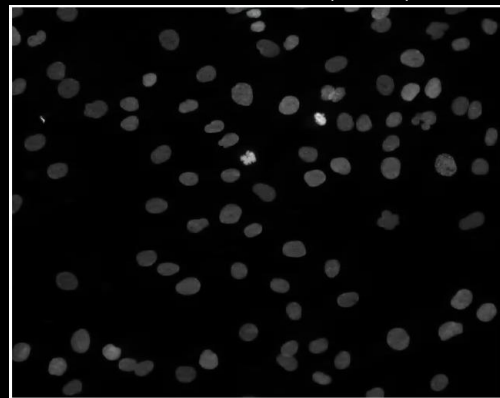
Extending UNet_{2L} For Sub-Cellular Structures

Chr Paint Mask (KNIP)



Chr.18

TERT-CRL-1474: DAPI , MIP, Bin-2



Acquisition: 300-500 ms

Results: 500-800 ms vs 2-4 s (seeded Watershed)

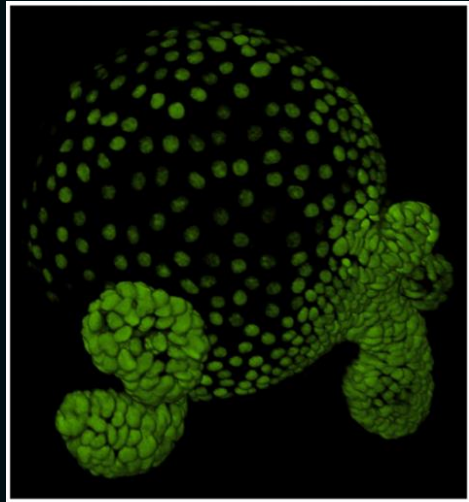
Image Size: 1276x1076 pixels

Summary

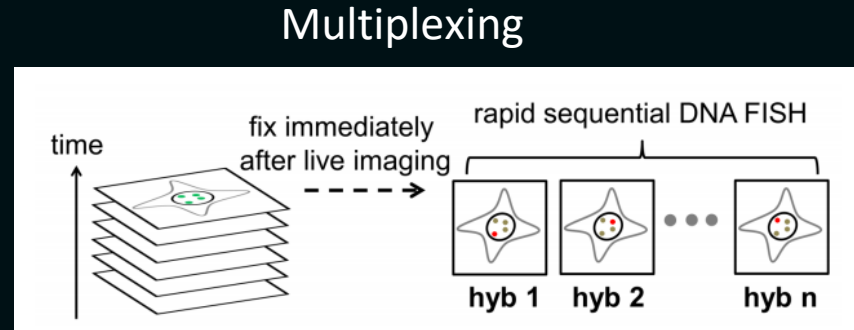
- Rapid prototyping for quantitative bio-imaging
 - Image Processing + ML + DL + Python + R
- Assess your requirements
 - KNIME forum(s), KNIP on GitHub
 - Hardware: SSD and RAM
 - KNIME Server (commercial)

Future Work

- KNIME Server/WebPortal
 - Storage integration with KNIME
- KNIME + ~~Deep Learning 4 Java~~ (Tensorflow-JNI)



Clevers, Cell, 2013



Source: Guan et. al., Biophysical Journal doi: 10.1016/j.bpj.2017.01.032

Acknowledgements

HiTIF

G. Pegoraro

L. Ozbun

A. Carcamo

Misteli Lab

K. Nakayama

Z. Jowhar

S. Shachar

T. Misteli

Larson Lab

D. Larson

Y. Wan

J. Roderiguez

M. Palangat

D. Larson

Hager Lab

G. Hager

D. Stavreva

Others

G. Zaki, HPC@CBIIT

CBIIT Server Team

Biowulf, HPC@CIT

ABCC, FNLRC

KNIME + KNIP Team

**Partially Funded By the National Cancer Institute, National Institutes of Health,
Project Num. 1ZICBC011567-01.**

Thank You

<https://github.com/CBIT/Misteli-Lab-CCR-NCI>

“The content of this presentation does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.”

Contact Information

Prabhakar R. Gudla

gudlap@mail.nih.gov

<https://github.com/CBIIT/Misteli-Lab-CCR-NCI>

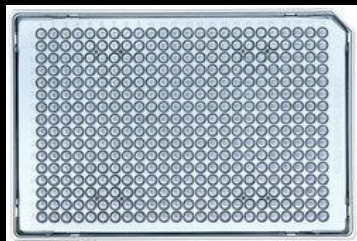
High-Throughput Imaging Facility (HiTIF)

Laboratory of Receptor Biology and Gene Expression

Center for Cancer Research   NATIONAL CANCER INSTITUTE

Robustness of UNet_{2L}

Fibroblast (Normoxia)
2 FISH Spots/ Cell



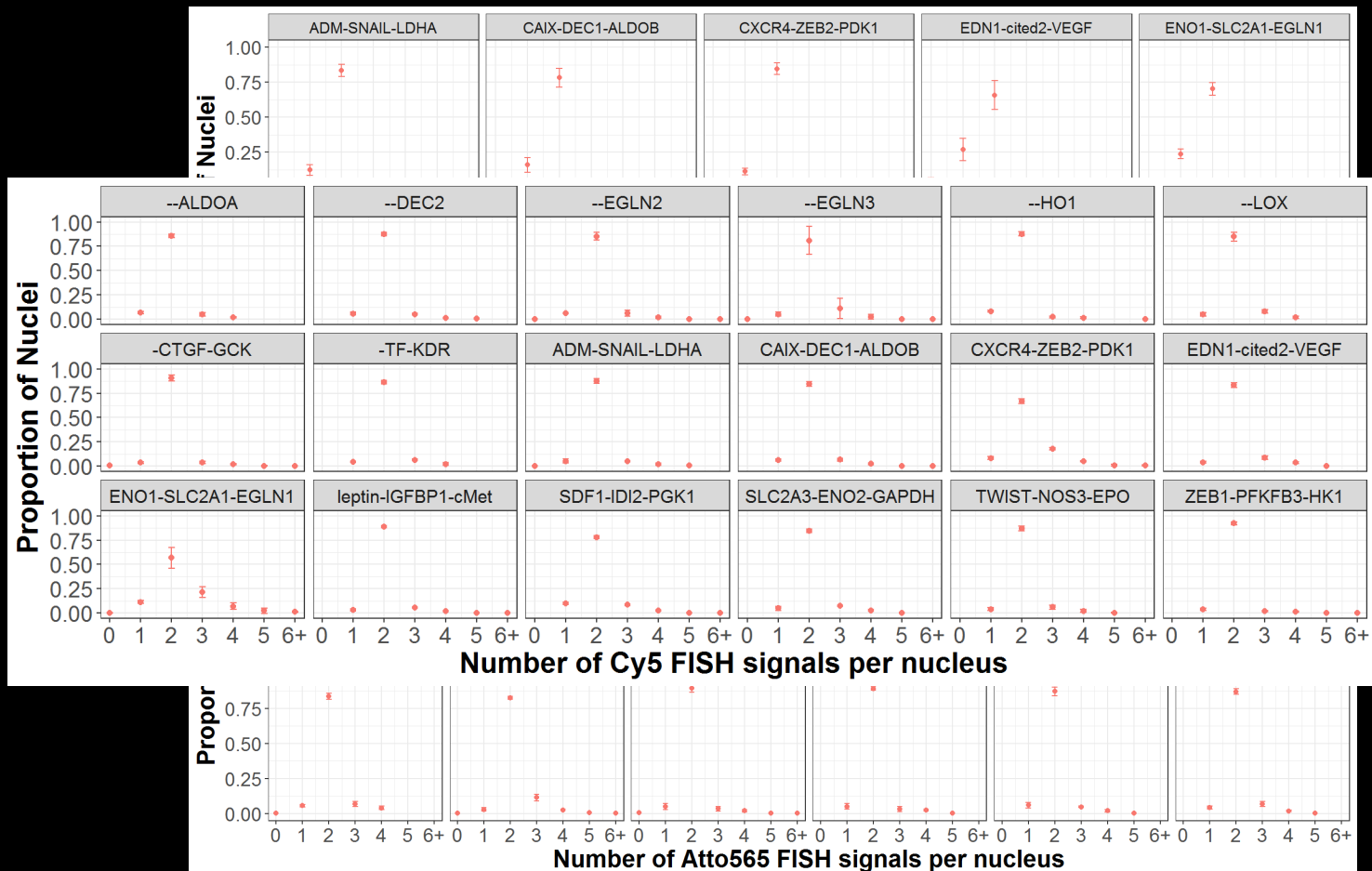
Test-P1



Quadruplicates

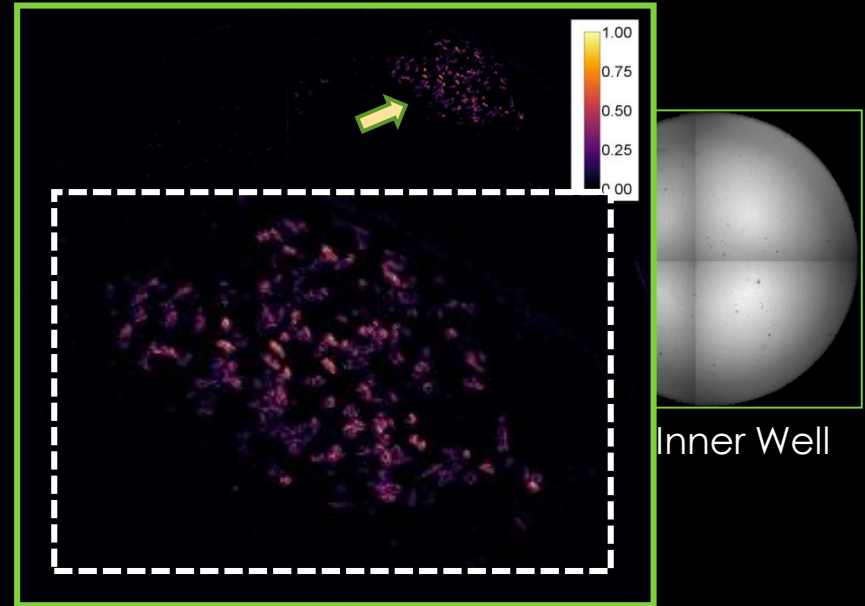
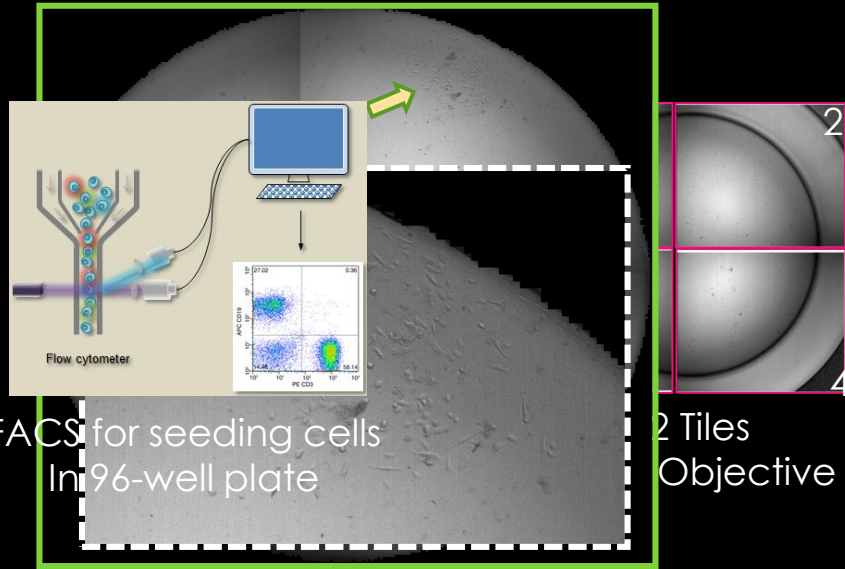


>0.99



Finding AOIs in Plate

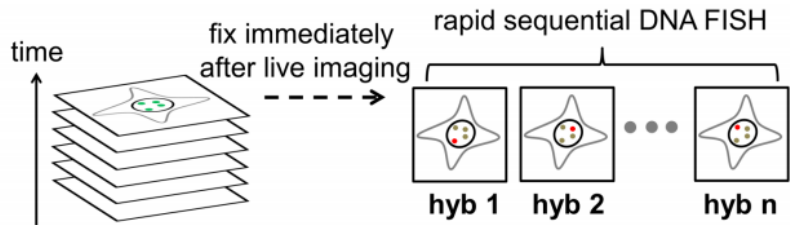
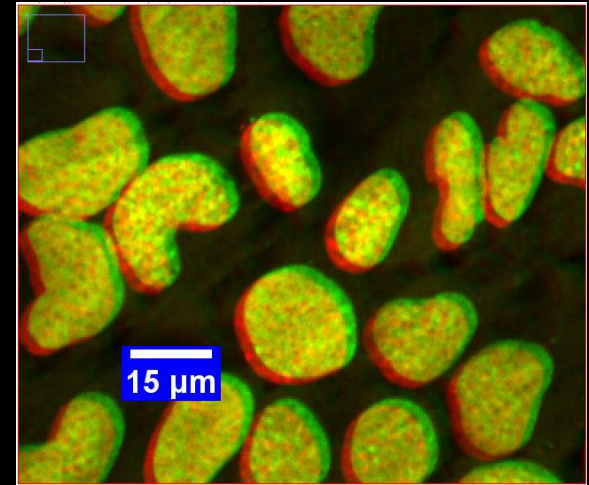
- First Pass: Brightfield, low magnification objective (e.g., 4X, Dry)
- Find Area(s) of Interest (AOI) using pixel-level segmentation
- Second Pass: Go back to AOI and image with higher magnification



Sequential DNA FISH

CV7000, Objective: 40X, Dry

Before (Red: Run-1, Green: Run-2)



Source: Guan et al., *Biophysical Journal* doi: [10.1016/j.bpj.2017.01.032](https://doi.org/10.1016/j.bpj.2017.01.032)

After Registration (Yellow is better)

