7th KNIME User Group Meeting

Automated light microscopy -based imaging and data analysis using KNIME Image Processing

RNAi Screening Facility
BioQuant, Heidelberg

Jürgen Reymann
• **Scientific Background**  
  (...from a data analysts perspective)

• **Data Analysis**

• **Automated Imaging**

  Platform independant Solution
  Fully Integrated
BioQuant
Quantitative Analysis of Molecular and Cellular Biosystems

Interaction of viral and cellular Systems

Modeling Platform

Technology Platform

Large Scale Data Facility

ViroQuant-CellNetworks
RNAi Screening Facility

NIKON IMAGING CENTER

13/02/2014

KNIME User Group Meeting
RNAi Screening Facility

Screening Core Facility
Assay automation | Automated microscopy

Data Analysis
Automated data processing pipelines

Research & Development
Novel technologies for HCS
RNAi screening platform to identify host factors involved in HIV-1 (HCV, DV, AAV, PV) replication

Screening Core Facility
Assay automation | Automated microscopy

Information / Data Pipeline

Data Analysis
Automated data processing pipelines

Transfection protocol
- siRNA library
- Assay automation
- Plate layouts

Sample preparation
- Screen meta data

Data acquisition
- Screen meta data
- High-throughput
- High-content
- Correlative
- Data storage
- Image raw data
- Image meta data

Image Processing
- Analysis results
- Quality control
- Normalisation
- Segmentation
- Features

Statistics
- Analysis results
- Normalisation (per plate/assay)
- Spatial normalisation
- Hit calling

Bioinformatics
- Analysis results
- Data integration
- Processes
- Pathways
- Interaction networks

Visualisation (plate viewer, image viewer, plots) | Tables | Meta data | … |

Data Base
Information / Data Pipeline

Data Analysis
Automated data processing pipelines

Image Reader

Experiment MetaData
Libraries
Layouts

Image Processing

Statistics
Bioinformatics
Database Writer

Results
Information / Data Pipeline

96er head pipetting robot

Well plates

384 source plate

→ Transformation in 384 well plate format (source plate)

384 cell arrays (LabTeks)

Whole genome cell arrays

Biological Assays

KNIME User Group Meeting

13/02/2014

6/38er head printing robot

384er head printing robot

96er head pipetting robot

Well plates

384 source plate

→ Transformation in 384 well plate format (source plate)

384 cell arrays (LabTeks)

Whole genome cell arrays
3x Olympus IX81 ScanR

2x (exp.) Leica TCS SP5

Perkin Elmer Opera LX
• Scientific Background  
  (...from a data analysts perspective)

• Data Analysis

• Automated Imaging

Platform independant Solution
Fully Integrated

KNIME  Leica
3D high-resolution data stacks

- Nuclei
- Promyelocytic leukemia (PML) nuclear bodies
- Telomeres

► Colocalisation  PML ↔ Telomeres

experimental Leica TCS SP5
Image Processing

Use Case I :: Colocalisation
Image Processing

Use Case I :: Colocalisation
Image Processing

Use Case II :: n-dim Feature Space Analysis

2D high-throughput images

- Nuclei (NO phenotypic penetration)
Image Processing

Use Case II :: n-dim Feature Space Analysis

2D high-throughput images

- Nuclei (NO phenotypic penetration)
- siRNA INCENP phenotypes
Image Processing

Use Case II :: n-dim Feature Space Analysis

Rel. distance of object to INCENP feature space

Positive classified (Rel. to Incenp feature space)

Negative classified (Rel. to Incenp feature space)

n = 8

n = 32

Erb

n = 35

KIF11

n = 8

INCENP

n = 35

PLK

Object
- Scientific Background  
  \((\text{...from a data analysts perspective})\)

- Data Analysis

- Automated Imaging
  Platform independant Solution
  Fully Integrated  
  \(\text{Leica KNIME}\)
n-dim feature space of observables
n-dimensional feature space of observables

- Feature 1
- Feature 2
- Feature n

Different characteristics for each object in n-dimensional feature space

**PROBLEM:** General measurement provides cloud of data points containing every feature, i.e., every object!

**BUT:** Every measurement aims at collecting the information content of specified objects
n-dimensional feature space of observables

General Considerations

FeedBack-driven DAQ

Different characteristics for each object in n-dimensional feature space

**PROBLEM:** General measurement provides a cloud of data points containing every feature, i.e., every object!

**BUT:** Every measurement aims at collecting the information content of specified objects.

**IDEA:**

- *Pre-data analysis* (identify objects of interest, e.g., *Feature 1*)
- FeedBack to measuring system
- Measure only objects of interest (No junk data!)

*Feature 1*, *Feature 2*, *Feature n*
General Considerations

**n-dim feature space of observables** [cells]

- Feature 1
- Feature 2
- Feature n

**FeedBack**

- FeedBack to measuring system
- Measure only objects of interest (No junk data!)

---

**High-resolution 3D colocalisation**

RNA interference screen

- 20 cell arrays
- 38,400 3D data stacks
- ~2 Mio images

Pure screening time: 6 months
Junk data: ~40%

Osterwald S. et al, Biotechnology J. 2012
**General Considerations**

n-dim feature space of observables [cells]

- Feature n
- Intensity

**Colocalisations**
- objects of interest

**High-resolution 3D colocalisation**

RNA interference screen

- 20 cell arrays
- 38,400 3D data stacks
- ~2 Mio images

Pure screening time: 6 months
Junk data: ~40%

**FeedBack**

- FeedBack to measuring system
- Measure only objects of interest (No junk data!)

Osterwald S. et al, Biotechnology J. 2012
Automated Imaging

Combine/Link DAQ ↔ IP

- **Transfection protocol**
  - siRNA library

- **Sample preparation**
  - Assay automation
  - Plate layouts

- **Data acquisition**
  - High-throughput
  - High-content
  - Correlative
  - Data storage

- **Image Processing**
  - Quality control
  - Normalisation
  - Segmentation
  - Features

- **Statistics**
  - Normalisation (per plate/assay)
  - Spatial normalisation
  - Hit calling

- **Bioinformatics**
  - Data integration
  - Processes
  - Pathways
  - Interaction networks

Use information from Image Pre-Processing in order to optimise / upgrade DAQ
Automated Imaging

Combine/Link DAQ ↔ IP

Transfection protocol
siRNA library

Sample preparation
Assay automation
Plate layouts

Data acquisition
High-throughput
High-content
Correlative
Data storage

Image Processing
Quality control
Normalisation
Segmentation
Features

Statistics
Normalisation
(Sample normalisation)
Spatial normalisation
Hit calling

Bioinformatics
Data integration
Processes
Pathways
Interaction networks

Use information from Image Pre-Processing in order to optimise / upgrade DAQ

Junk data | Costs | Feasibility of experiments | Bleaching

- Trigger (time lapse experiments)
- Pick only objects of interest ➤ Quality / Information content of image raw data
  - Screening time
- Pick representative cells for high/super -resolution
- ...

Automated Imaging

13/02/2014
Automated Imaging

Data acquisition

Image Pre-Processing
- Targets | Trigger
- Hit calling

Image Processing
- Quality control
- Normalisation
- Segmentation
- Features
Automated Imaging

Data acquisition

Image Pre-Processing
- Targets | Trigger
- Hit calling

Objects of interest

Image Processing

Objects of interest

Image Reader
- Microscope Raw Data

Experiment MetaData
- Libraries
- Layouts

Image Processing

Microscope Feedback
- Targets
• **Scientific Background**  
  \( \text{(...from a data analysts perspective)} \)

• **Data Analysis**

• **Automated Imaging**

  **Platform independant Solution**

  Fully Integrated
**Idea**: Acquire images of the same sample/target at multiple systems, in order to combine the advantages of different microscopic techniques.

**Examples:**

- **High throughput data** select targets **High/Super-resolution imaging**
- **High throughput data** quality control **Complete missing data**
- **Low resolution 2D complete overview** expert pick **Super-resolution 3D imaging**
**Problem:** Transfer System

Coordinate transfer by

1. Reference markers
2. (SIFT) image registration
Automated Imaging I

Problem: Transfer System

Coordinate transfer by
1.) Reference markers
2.) (SIFT) image registration

System A  System B  System C

getRefPosLeicaSP5
getRefPosOlympusDX81
rotatePlateToTable
setPosSP5
rotateTableToPlate
getCellPos
rotatePlateToTable
setPosSTORM
getRefPosSTORM

High-resolution Leica SP5
10 µm

Super-resolution dSTORM
1 µm

Image Processing
Target identification

dSTORM setup: Mike Heilemann
High-throughput screening
[ → Target identification ]

Collab.: Mike Heilemann, Vytaute Starkuviene

Flottmann et al., Biotechniques 2013
• Scientific Background (…from a data analyst's perspective)

• Data Analysis

• Automated Imaging

Platform independent Solution

Fully Integrated
Automated Imaging II

experimental Leica TCS SP5

Data storage

(Pre-Screen)
Image raw data

Image Reader

Microscope Raw Data

Experiment MetaData

Libraries Layouts ...

Image Processing

Targets / Trigger

CAM Loop Start
CAM Path Parser
Image Pre-Processing
CAM Loop End

Communication with CAM server
Parse image paths
String command generator
Communication with CAM server

Collab.: KNIME

13/02/2014 KNIME User Group Meeting
Leica LASAF software - *Matrix Screener*

→ Experiment setup (job)
Automated Imaging II

Currently Available CAM Commands

### Start Scan

**Command to sent** /cli:frank /app:matrix /cmd:startscan

**Token Description:**
- /cli: = Client (the name of the client send this command)
- /app: = Application (the target application for this command)
- /cmd: = the command to proceed

**CAM Command Description:**
Will start the MatrixScreener scan (not the CAM Scan Field Scan!)

### Stop Scan

**Command to sent** /cli:frank /app:matrix /cmd:stopscan

**Token Description:**
- /cli: = Client (the name of the client send this command)
- /app: = Application (the target application for this command)
- /cmd: = the command to proceed

**CAM Command Description:**
Will stop the MatrixScreener scan (not the CAM Scan Field Scan!)

### Pause Scan

**Command to sent** /cli:frank /app:matrix /cmd:pausescan

**Token Description:**
- /cli: = Client (the name of the client send this command)
- /app: = Application (the target application for this command)
- /cmd: = the command to proceed

**CAM Command Description:**
Will switch the MatrixScreener scan into the Pause modus

and lots more...!
**Example:**

1. Define **job1** ➤ Pre-Screen, e.g. **2D images** (focus plane) of full sample using **1 colour channel**

2. Define **job2** ➤ Target screen, e.g. **3D data stacks** of objects of interest using **3 colour channels**
1. **job1** ► Pre-Screen: 2D images (focus plane) of full sample using 3 colour channels

   **Image Pre-Processing** ► Find candidates for colocalisation
   
   [ Use Case I :: Colocalisation ]

2. **job2** ► Target screen: 3D data stacks of candidates using 3 colour channels

   Biological question requires **job2** as experiment setup to gain information!
Automated Imaging II

Test Setup :: KNIME Workflow

Collab.: KNIME

Communication with CAM server

CAM Loop Start

CAM Path Parser

parse image paths

CAM Loop End

Communication with CAM server

Java Snippet

Inject empty cmd

no scan image ->
(no image or a cam level image)

This branch completes the loop if we don't want to analyze an image.

Either it is no image or it is not part of the scan but an detail image (cam level)

In both cases an empty cmd is injected to close the loop.

Java IF (Table)

D = {picture || cam

End IF

Column Filter

command

End IF

Java Snippet

Conditions for CAM-Feedback

Build response command

Column Filter

command
**Pre-Screen**

*job1 ➤ Pre-Screen: 2D images (focus plane) of full sample*
Automated Imaging II

Test Setup :: KNIME Workflow

Pre-Screen

CAM

Dialog - 296 - CAM Loop Start (Communication with)

File

Options   Advanced   Flow Variables   Memory Policy
connection

server: 127.0.0.1
port: 8895

settings:
start command: /knime/appmatrix/cmnd/startscan

OK   Apply   Cancel

Dialog - 297 - CAM Path Parser (parse image)

File

Options   Advanced   Flow Variables   Memory Policy

root folder of microscope images

Selected Directory:

C:\MicroscopeImages

Browse...

CAM command column: $CamCommand

OK   Apply   Cancel

Collab.: KNIME
Automated Imaging II

Test Setup :: KNIME Workflow

Pre-Screen

Table "default" - Rows: 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 0</td>
<td>Image 0000-0000-0000-0000-0000-0000-0000-0000</td>
<td>Nucleus</td>
<td>Nucleus_119</td>
<td>439.851</td>
<td>390.905</td>
<td>6</td>
<td>00</td>
<td>51</td>
<td>00</td>
<td>00</td>
<td>01</td>
<td>440</td>
<td>391</td>
</tr>
<tr>
<td>Row 1</td>
<td>Image 0000-0000-0000-0000-0000-0000-0000-0000</td>
<td>Nucleus</td>
<td>Nucleus_07</td>
<td>325.37</td>
<td>240.301</td>
<td>7</td>
<td>00</td>
<td>51</td>
<td>00</td>
<td>00</td>
<td>01</td>
<td>325</td>
<td>240</td>
</tr>
</tbody>
</table>

read the scan image and (pre-) process...

Communication with CAM server

Condition for CAM-Feeback

(> = 1 Nucleus/Imag with > = 5 COLOC's)

Collab.:
Automated Imaging II

Test Setup :: KNIME Workflow

Pre-Screen

CAM Loop Start

CAM Path Process

Java IF (Table)

End IF

CAM Loop End

Communication with CAM server

Table "default" - Rows: 2

Spec: Columns: 3 | Properties | Flow variables

Raw ID | Object | Label | ID | ORCID

Row 1: Row 0: Nucleus | Nucleus_119

Row 2: Row 0: Nucleus | Nucleus_57

Communication with CAM server

Pre-Screen

Image Reader

Image Pre-Processing

Java Snippet

Column Filter

Conditions for CAM-Feedback

build response command

no scan image -> (no image or a cam level image)

This branch completes the loop if we don't want to analyze an image.

Either it is no image or it is not part of the scan but an image (cam level)

In both cases an empty cmd is injected to close the loop.

Collab.: KNIME Leica
Target Screen

**job2**  ➤ Target screen: 3D data stacks of candidates *(Nucleus_119 | Nucleus_67)*
**Automated Imaging II**  

**Test Setup :: Results**

**Highly accurate positioning of targets in the center of field of views**

<table>
<thead>
<tr>
<th>Row 0</th>
<th>Row 1</th>
<th>Row 2</th>
<th>Row 3</th>
<th>Row 4</th>
<th>Row 5</th>
<th>Row 6</th>
<th>Row 7</th>
<th>Row 8</th>
<th>Row 9</th>
<th>Row 10</th>
<th>Row 11</th>
<th>Row 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row 13</td>
<td>Row 14</td>
<td>Row 15</td>
<td>Row 16</td>
<td>Row 17</td>
<td>Row 18</td>
<td>Row 19</td>
<td>Row 20</td>
<td>Row 21</td>
<td>Row 22</td>
<td>Row 23</td>
<td>Row 24</td>
<td>Row 25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row 26</td>
<td>Row 27</td>
<td>Row 28</td>
<td>Row 29</td>
<td>Row 30</td>
<td>Row 31</td>
<td>Row 32</td>
<td>Row 33</td>
<td>Row 34</td>
<td>Row 35</td>
<td>Row 36</td>
<td>Row 37</td>
<td>Row 38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row 39</td>
<td>Row 40</td>
<td>Row 41</td>
<td>Row 42</td>
<td>Row 43</td>
<td>Row 44</td>
<td>Row 45</td>
<td>Row 46</td>
<td>Row 47</td>
<td>Row 48</td>
<td>Row 49</td>
<td>Row 50</td>
<td>Row 51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row 52</td>
<td>Row 53</td>
<td>Row 54</td>
<td>Row 55</td>
<td>Row 56</td>
<td>Row 57</td>
<td>Row 58</td>
<td>Row 59</td>
<td>Row 60</td>
<td>Row 61</td>
<td>Row 62</td>
<td>Row 63</td>
<td>Row 64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row 65</td>
<td>Row 66</td>
<td>Row 67</td>
<td>Row 68</td>
<td>Row 69</td>
<td>Row 70</td>
<td>Row 71</td>
<td>Row 72</td>
<td>Row 73</td>
<td>Row 74</td>
<td>Row 75</td>
<td>Row 76</td>
<td>Row 77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row 78</td>
<td>Row 79</td>
<td>Row 80</td>
<td>Row 81</td>
<td>Row 82</td>
<td>Row 83</td>
<td>Row 84</td>
<td>Row 85</td>
<td>Row 86</td>
<td>Row 87</td>
<td>Row 88</td>
<td>Row 89</td>
<td>Row 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
<td>Rows</td>
</tr>
</tbody>
</table>

**Experiment setup**

**Overall sample size:**  
8 * 12 = 96 positions

**Standard DAQ  ** job2

- Screening time: ~70h
- junk data / less targets (NOT centered !)

**Automated Imaging II Test Setup :: Results**

**Overall acquisition time:** 7 h 17 min

**Highly accurate positioning of targets in the center of field of views**
Automated Imaging II

Test Setup :: Results

Highly accurate positioning of targets in the center of field of views

Experiment setup

Overall sample size: 8 * 12 = 96 positions

Standard DAQ  job2

- Screening time: ~70h
- junk data / less targets (NOT centered!)

Automated Imaging

Pre-Screen  job1

Target Screen  job2

- Screening time: ~7h
- 146 target regions (in center of field of view!)

- Decreased overall screening time (factor 10)
- Increased number and quality (centered) of acquired target regions

Overall acquisition time: 7 h 17 min
Automated Imaging

Conclusion

Overall sample size: 96 positions

Standard DAQ \textit{job2}

» Screening time: \textasciitilde 70h

Automated Imaging

Pre-Screen \textit{job1}

Target Screen \textit{job2}

» Screening time: \textasciitilde 7h

Costs

Standard DAQ \textit{job2}

\$ \textasciitilde 2800\ €

Automated Imaging

Pre-Screen \textit{job1}

Target Screen \textit{job2}

\$ \textasciitilde 280 \ €

Costs confocal microscopy: 40 \ € per hour

- Decreased overall screening time (factor 10)
- Increased number and quality (centered) of acquired target regions

Overall sample size: 38,400 positions

Standard DAQ \textit{job2}

» Screening time: 6 months

» \textasciitilde 40\% junk data

Automated Imaging

Pre-Screen \textit{job1}

Target Screen \textit{job2}

» Screening time: 3,6 months

\$ \textasciitilde 170,000\ €

\$ \textasciitilde 104,000\ €

\textasciitilde 66,000\ € junk data
Knowledge of Biological Systems

Perspectives…

- Biological Question
- Assay Development
- Image / Data Analysis
- Microscopy
- Experiment Setup
Technological improvements

Biology
- Preparation Methods
- Labeling Strategies
- Parallel Assays
- ...

Microscopy
- Automated Imaging
- Correlative
- Throughput
- Resolution
- ...

Knowledge of Biological Systems

Perspectives…
Perspectives… Knowledge of Biological Systems

Technological improvements

Biology
- Preparation Methods
- Labeling Strategies
- Parallel Assays
- …

Microscopy
- Automated Imaging
- Correlative
- Throughput
- Resolution
- …

Image / Data Analysis

Assay Development

Experiment Setup

Biological Question

Optimise/Upgrade DAQ
Perspectives…

Knowledge of Biological Systems

Technological improvements

Biology
- Preparation Methods
- Labeling Strategies
- Parallel Assays
- …

Microscopy
- **Automated Imaging**
- Correlative
- Throughput
- Resolution
- …

Increase of Information Content  Extraction / Interpretation of Information
Technological improvements

Biology
- Preparation Methods
- Labeling Strategies
- Parallel Assays
- …

Microscopy
- Automated Imaging
  - Experiment Setup
- Correlative
- Throughput
- Resolution
- …

Increase of Information Content → Extraction / Interpretation of Information

KNIME User Group Meeting
Acknowledgements

Holger Erfle
Manuel Gunkel
Tautvydas Lisauskas
Ruben Bulkescher
Bastian Schumacher
Jürgen Beneke
Nina Beil

Michael Berthold
Thomas Gabriel
Christian Dietz
Martin Horn
Michael Zinsmeier

Constantin Kappel
Frank Sieckmann
Werner Knebel

Mike Heilemann
Benjamin Flottmann