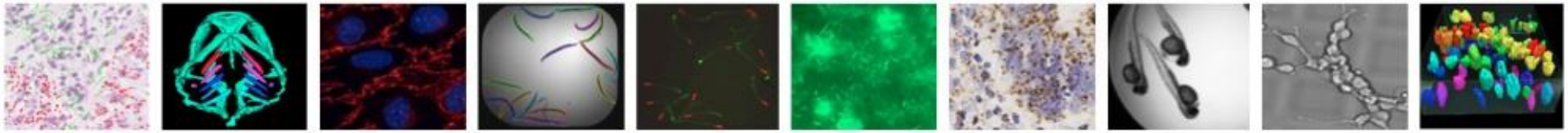


Analysis of microscopy data using

CellProfiler
cell image analysis software

CellProfiler™ Analyst
data exploration software

Carolina Wählby

Professor in Quantitative Microscopy
Centre for Image Analysis, Dept. of Information Technology, and SciLifeLab
Uppsala University, Uppsala, Sweden
and
Imaging Platform, Broad Institute of Harvard and MIT, Cambridge, MA, USA

BROAD INSTITUTE IMAGING PLATFORM SciLifeLab UPPSALA UNIVERSITET Centre for Image Analysis Uppsala University

Outline

- What is CellProfiler?
  - extracting measurements from cells
  - other functionality – Worm Toolbox and interfacing with Ilastik and Fiji/ImageJ

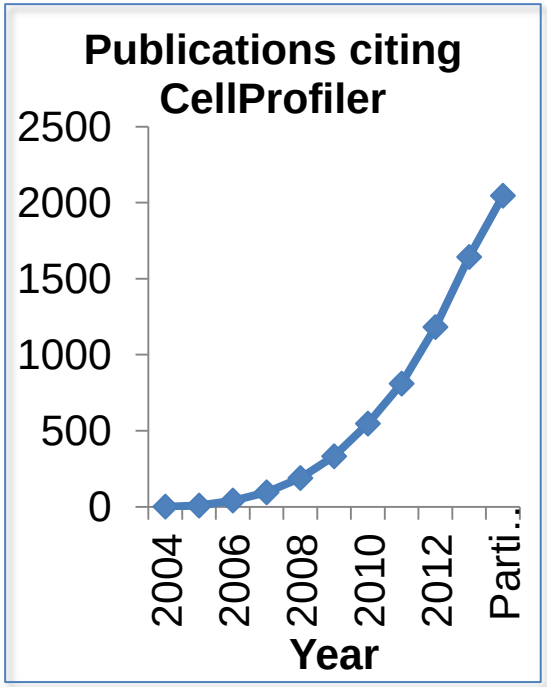
- What is CellProfiler Analyst?
  - exploring and visualizing extracted measurements
  - other functionality – Time-Lapse Tool, multi-collaborator data access and exploration

- CellProfiler and CellProfiler Analyst as tools for helping others get started with their own analysis within SciLifeLab Sweden

CellProfiler

cell image analysis software

- Published in 2006 (Carpenter et al, Genome Biology)
- Cited in **>2,000** scientific papers
- Launched **>100,000** times/year
- One of the **Top 10** most-accessed papers of all time in Genome Biology
- Winner of Bio-IT World's IT & Informatics Best Practices Award (2009)
- Interfaces with ImageJ/Fiji, ilastik, OMERO, KNIME, and more
- Originally based on Matlab, in 2009 completely re-written in Python
- www.cellprofiler.org

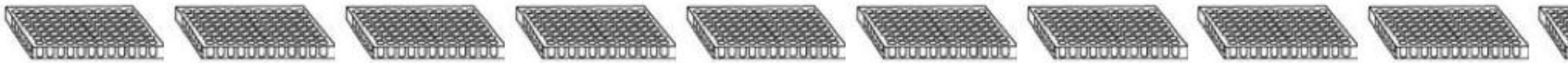


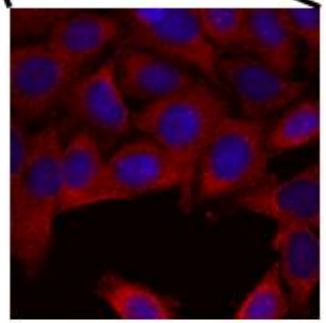

Anne Carpenter


Ray Jones

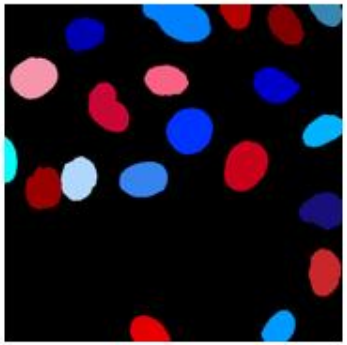

Lee Kamentsky

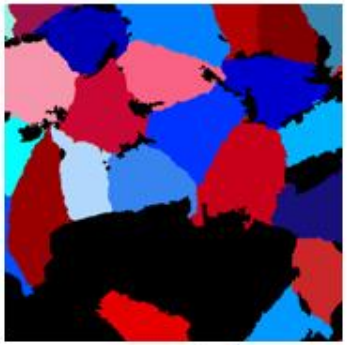
Cells or organisms in multiwell plates, each well treated with a gene or chemical perturbant





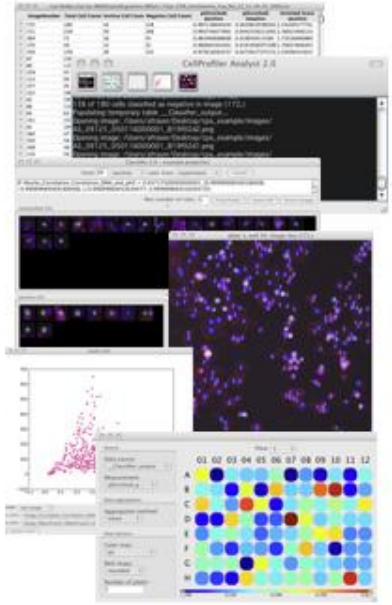
automated
microscopy
(any manufacturer)





Cell
measurements

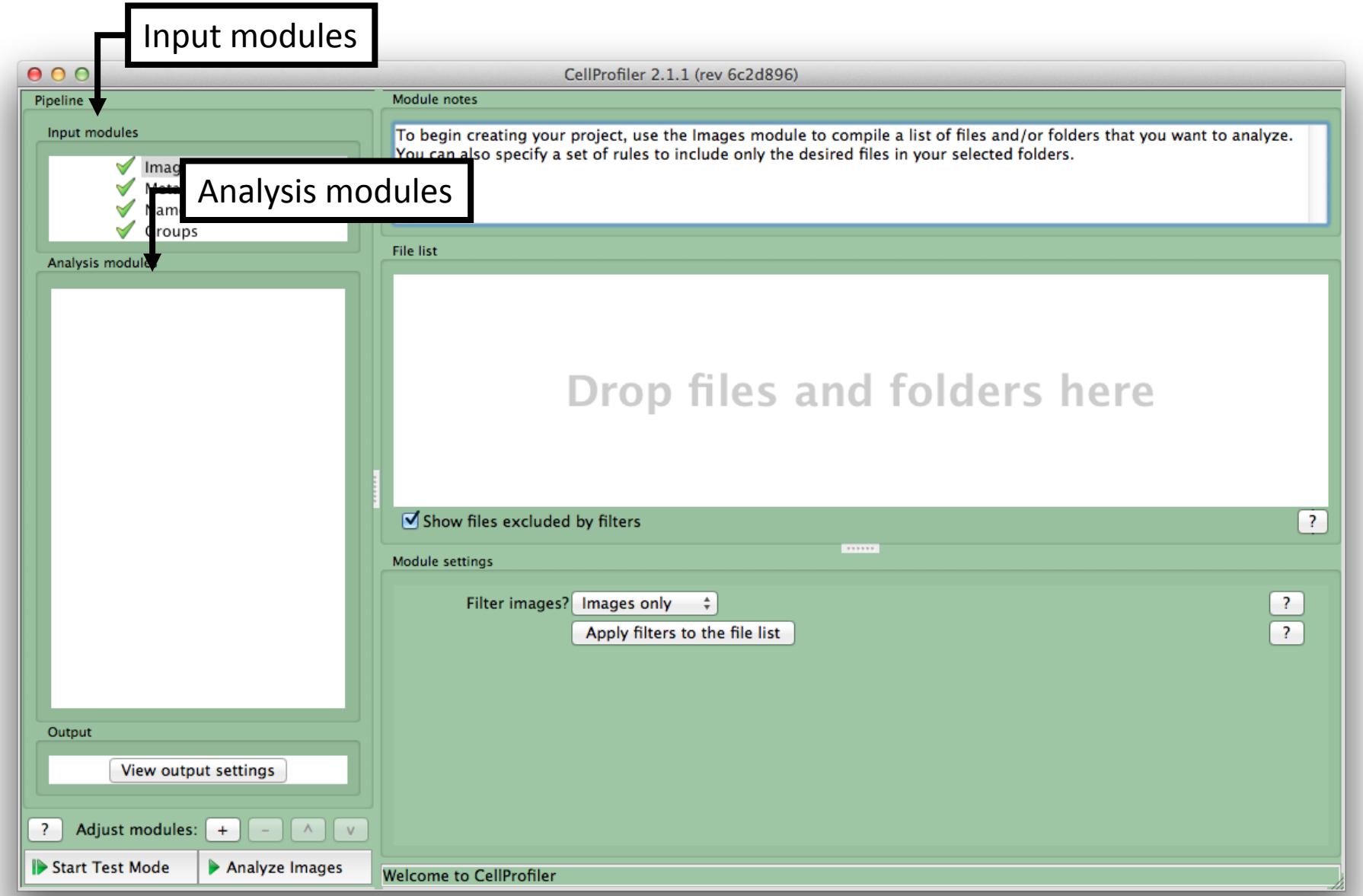
(size, shape,
intensity, texture,
etc.)



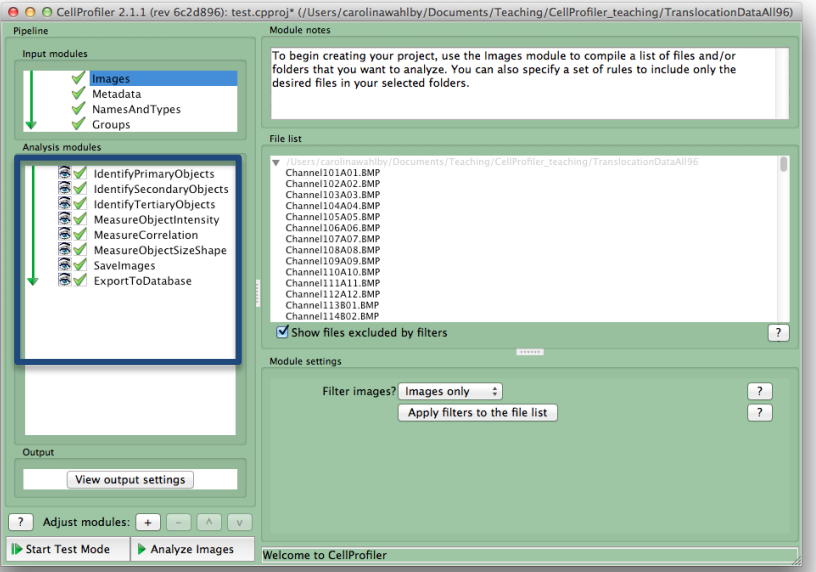
Data exploration
& machine learning

CellProfiler™
cell image analysis software

CellProfiler™
Analyst
data exploration software


Input modules
Analysis modules
CellProfiler 2.1.1 (rev 6c2d896)
Pipeline
Input modules
Groups
Analysis modules
Output
View output settings
Adjust modules:
Start Test Mode
Analyze Images
Module notes
To begin creating your project, use the Images module to compile a list of files and/or folders that you want to analyze.
You can also specify a set of rules to include only the desired files in your selected folders.
File list
Drop files and folders here
Show files excluded by filters
Module settings
Filter images?
Images only
Apply filters to the file list
Welcome to CellProfiler

Analysis Modules: setting up your pipeline

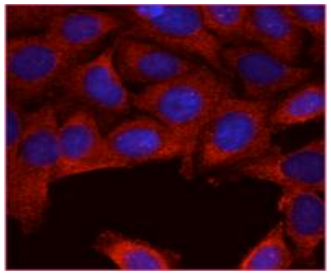


The analysis modules are built up based on a previous example pipeline or from scratch. Each step can be tested on a given or random image from the input modules, and the effects of the settings can be seen if the 'eye icon' is open.

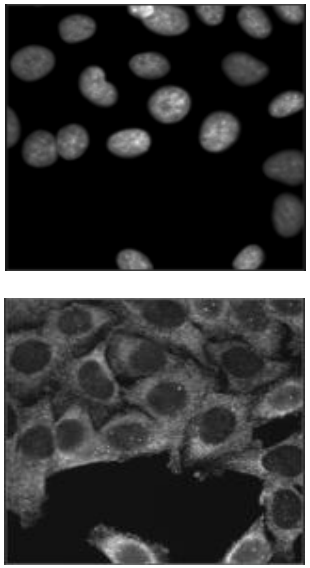
A wide range of specific as well as general-purpose functions are available.

Once the pipeline is completed, all images are analyzed by clicking 'Analyze images'

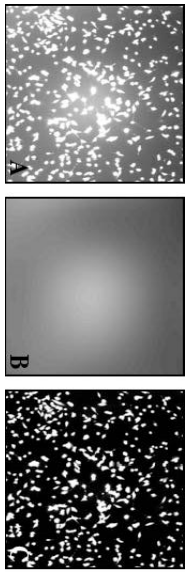
Original image



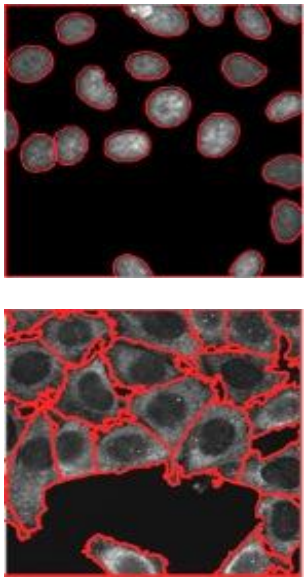
Split colors



Correct illumination



Identify cells/ compartments



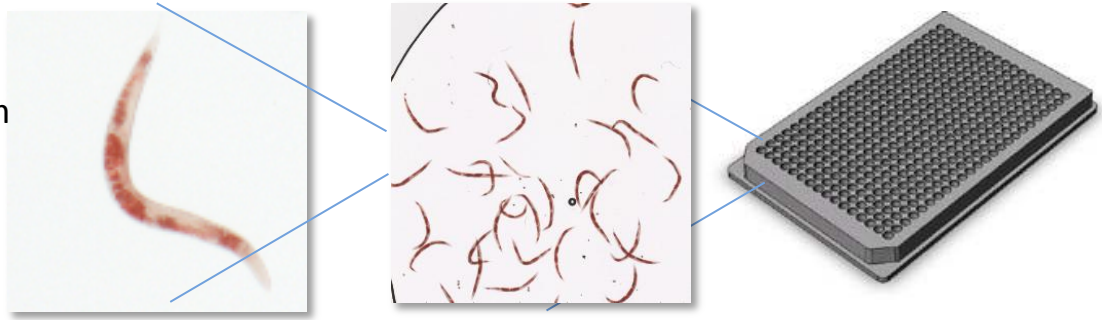
Measure 'everything'; Counts, Shapes, Sizes, Intensities, Textures, Correlations, Relationships …

Saved to a spreadsheet or database

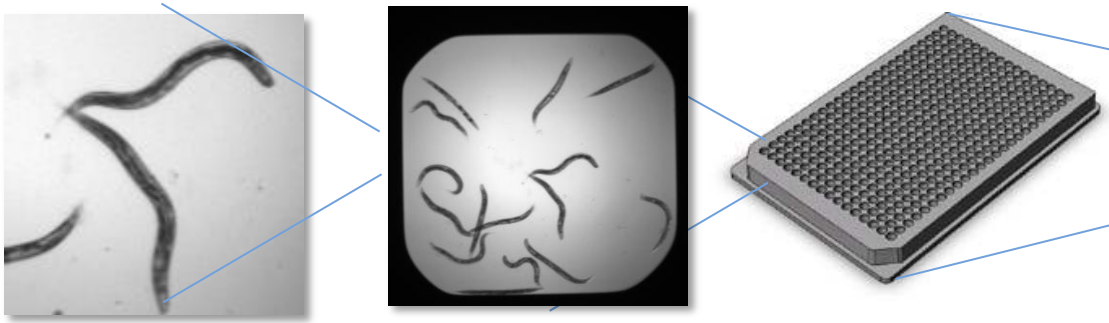
WormToolbox

C. elegans: a model organism used in 1600+ laboratories suitable for studying assays that can not be reduced to single cells:

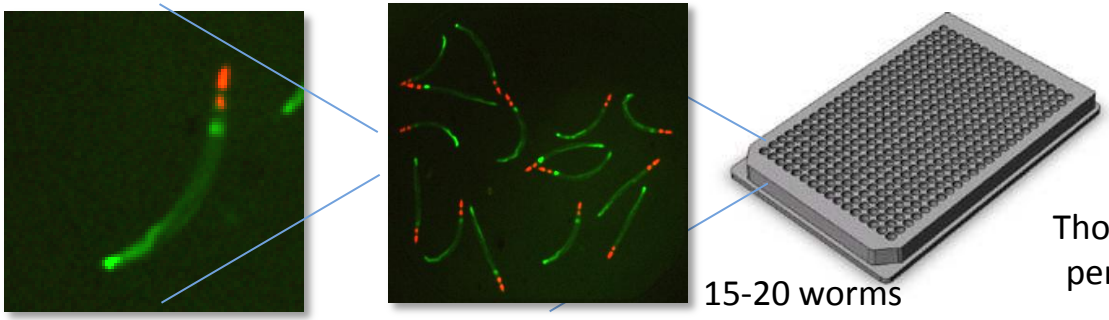
fat metabolism



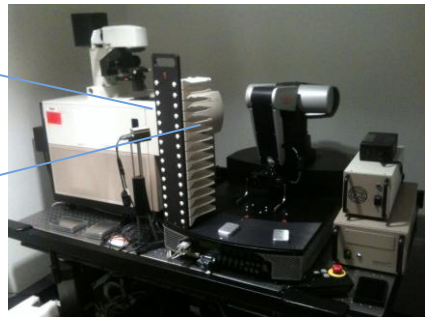
infectious disease



inter-tissue signaling



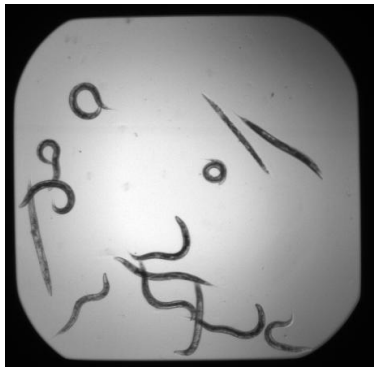
Stacks of multi-well plates can be imaged with automated microscopes



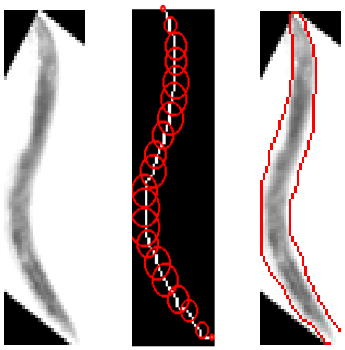
Millions of worms per multi-plate experiment

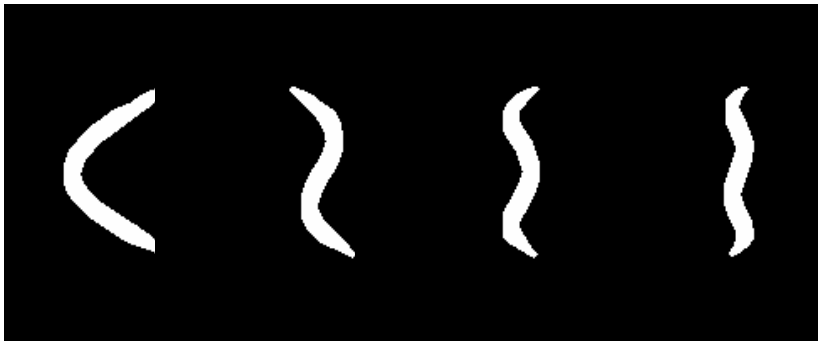
Thousands of worms per multi-well plate

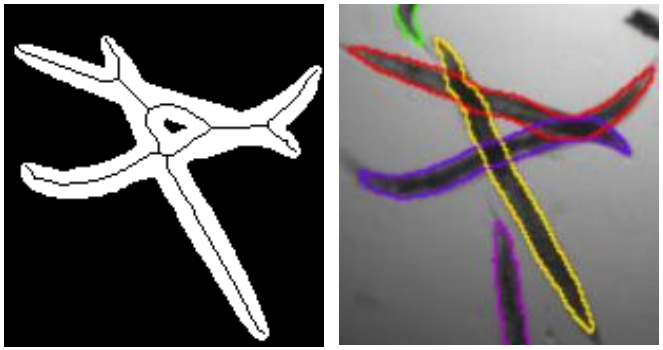
15-20 worms per well



Challenge; worms often overlap and cluster.
Approach: Find individual worms using a worm model.







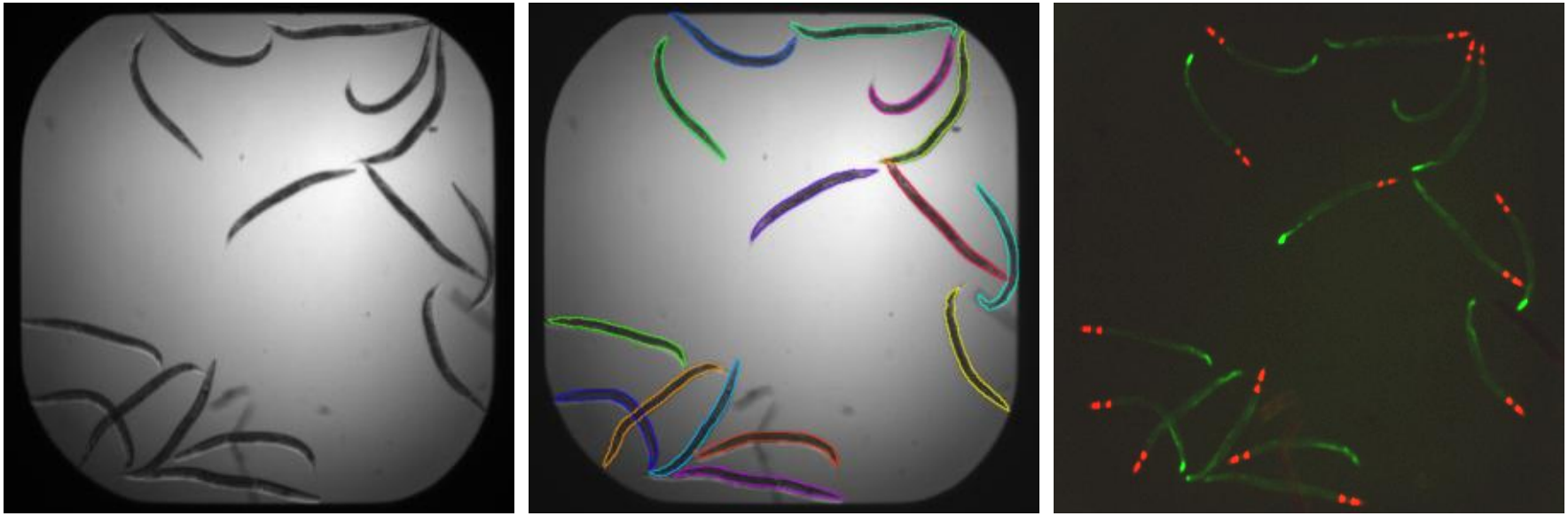
1. Select non-clustered 'training' worms and create a low-dimensional shape descriptor.

2. Combine many shape descriptors to a statistical model of worm shape.

3. Search for the best combination of worm shapes within each cluster.

Wählby et al ISBI 2010
Riklin-Raviv et al MICCAI 2010
Wählby et al Nature Methods 2012
Wählby et al Methods 2014

Once worms are untangled…

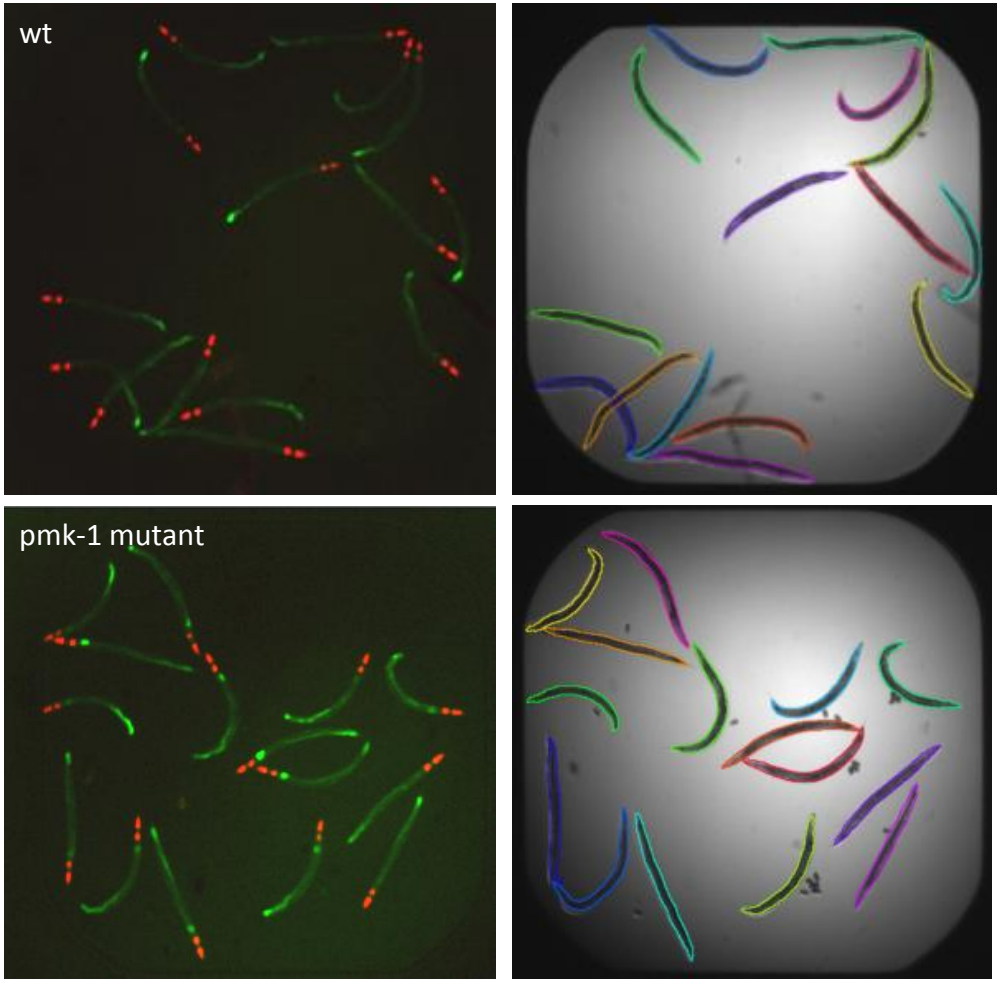


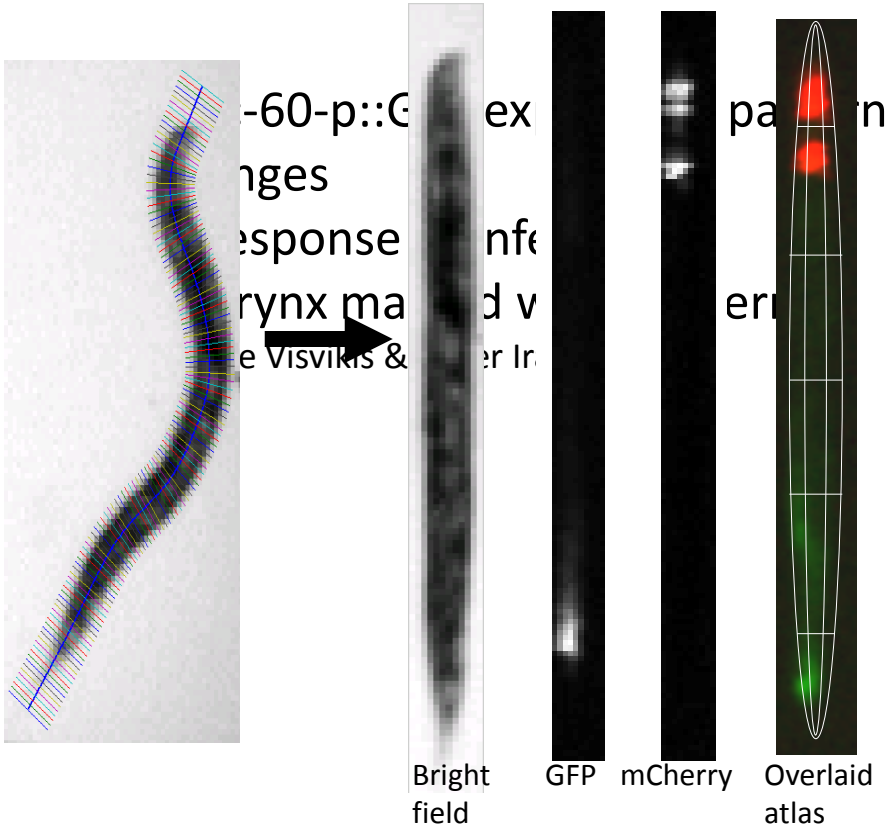
… we can

- measure size and shape
- count signals per worm
- quantify reporter signal patterns
- etc

… and increase throughput, objectivity and accuracy of phenotype scoring.

Quantification of variations in reporter protein expression



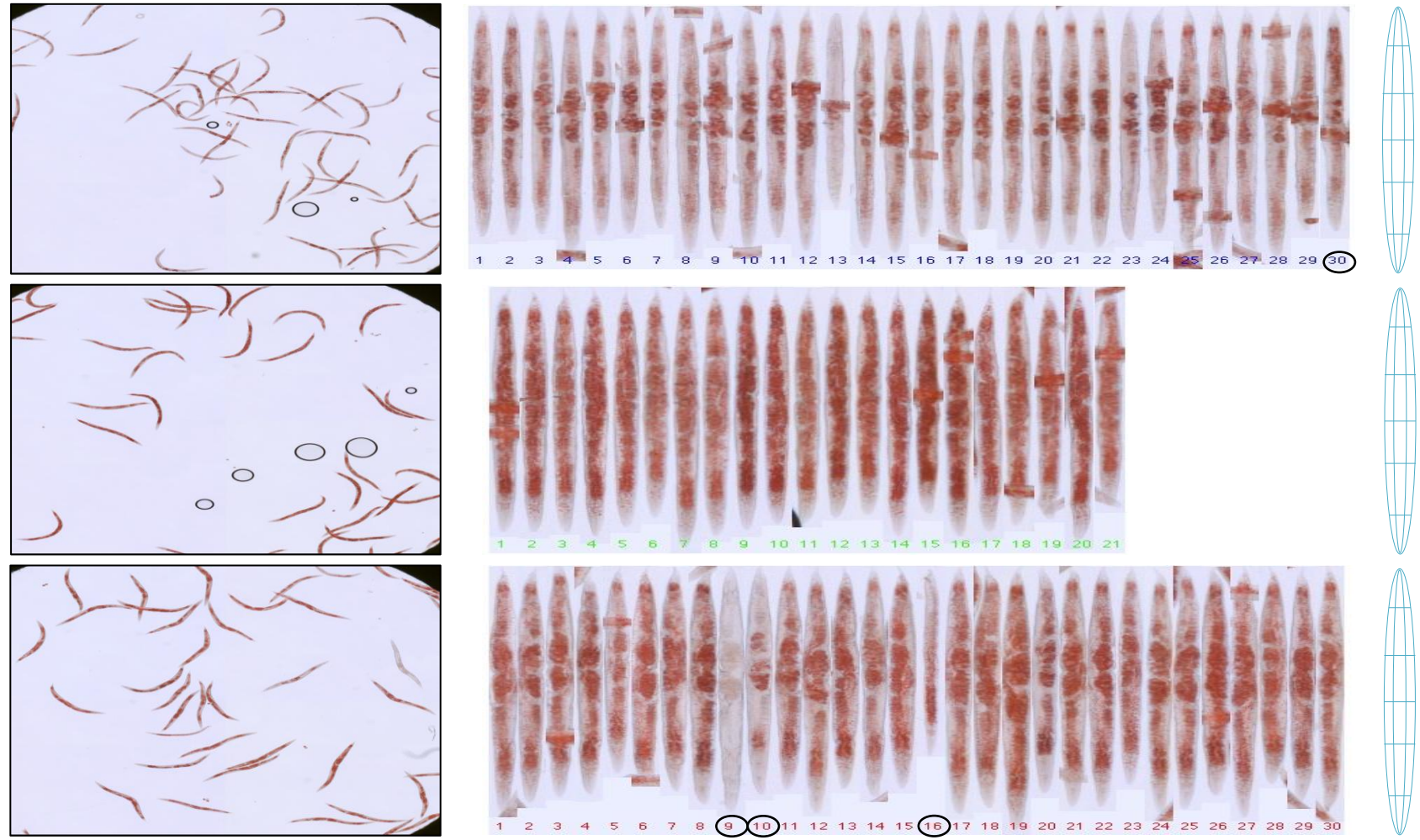


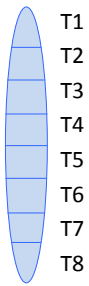
Once worms are delineated, we can resample each worm and align it with a common low-resolution atlas.

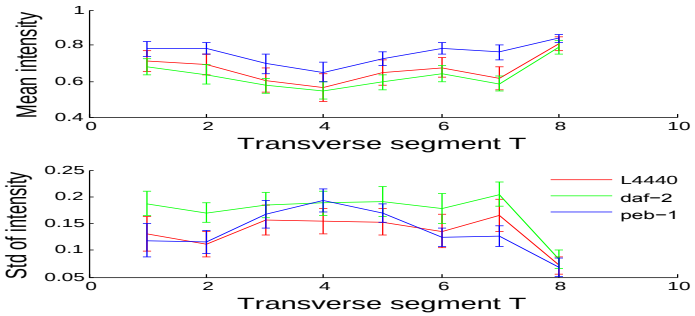
peb-1

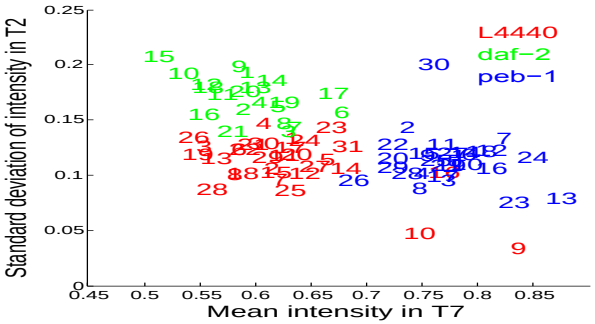
daf-2

L4440

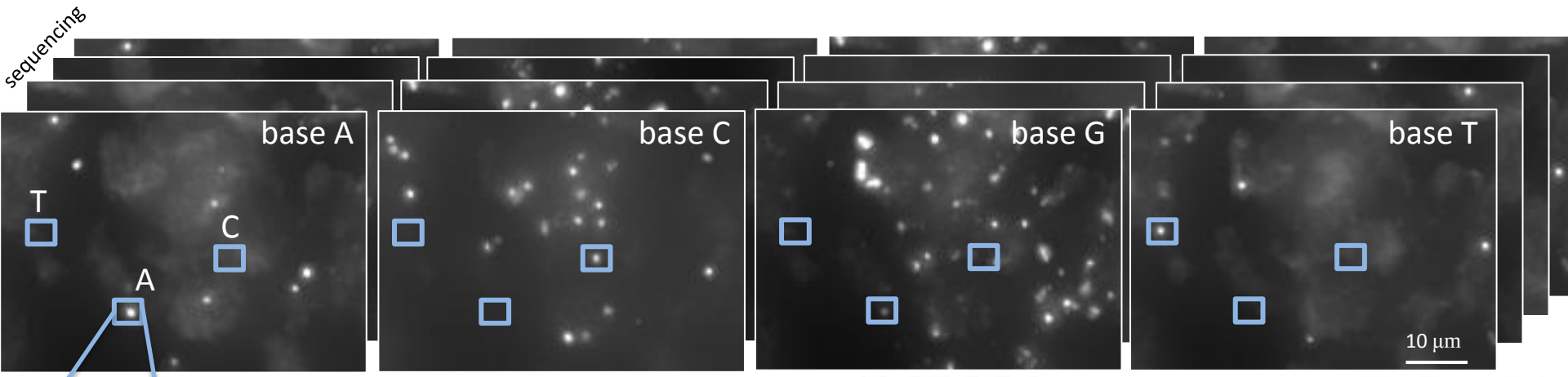


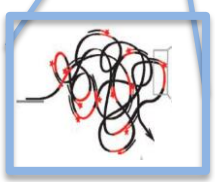






Interoperability CellProfiler cell image analysis software ⇔ ImageJ



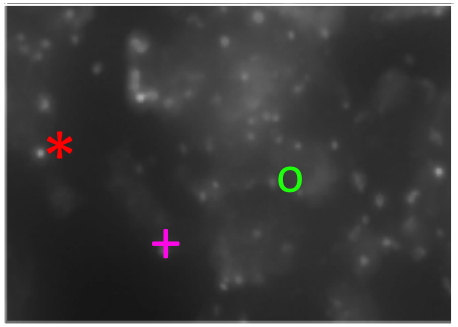


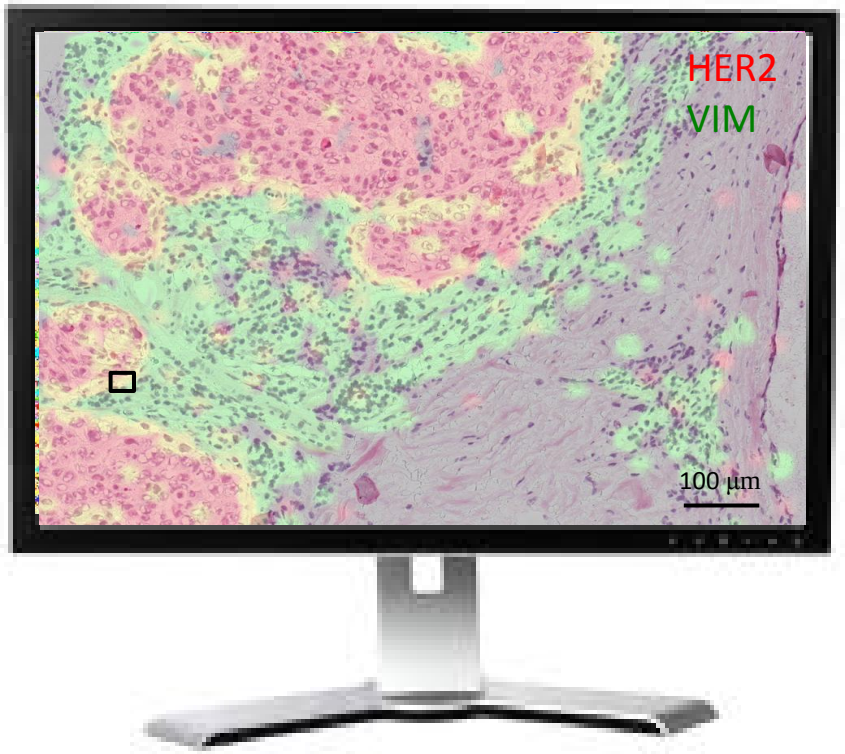
Detection and amplification

R. Ke, ..., **C. Wählby** and M. Nilsson, *Nature Methods*, 2013
A. Pacureanu,..., **C. Wählby**, *IEEE ISBI*, 2014

Sequence decoding

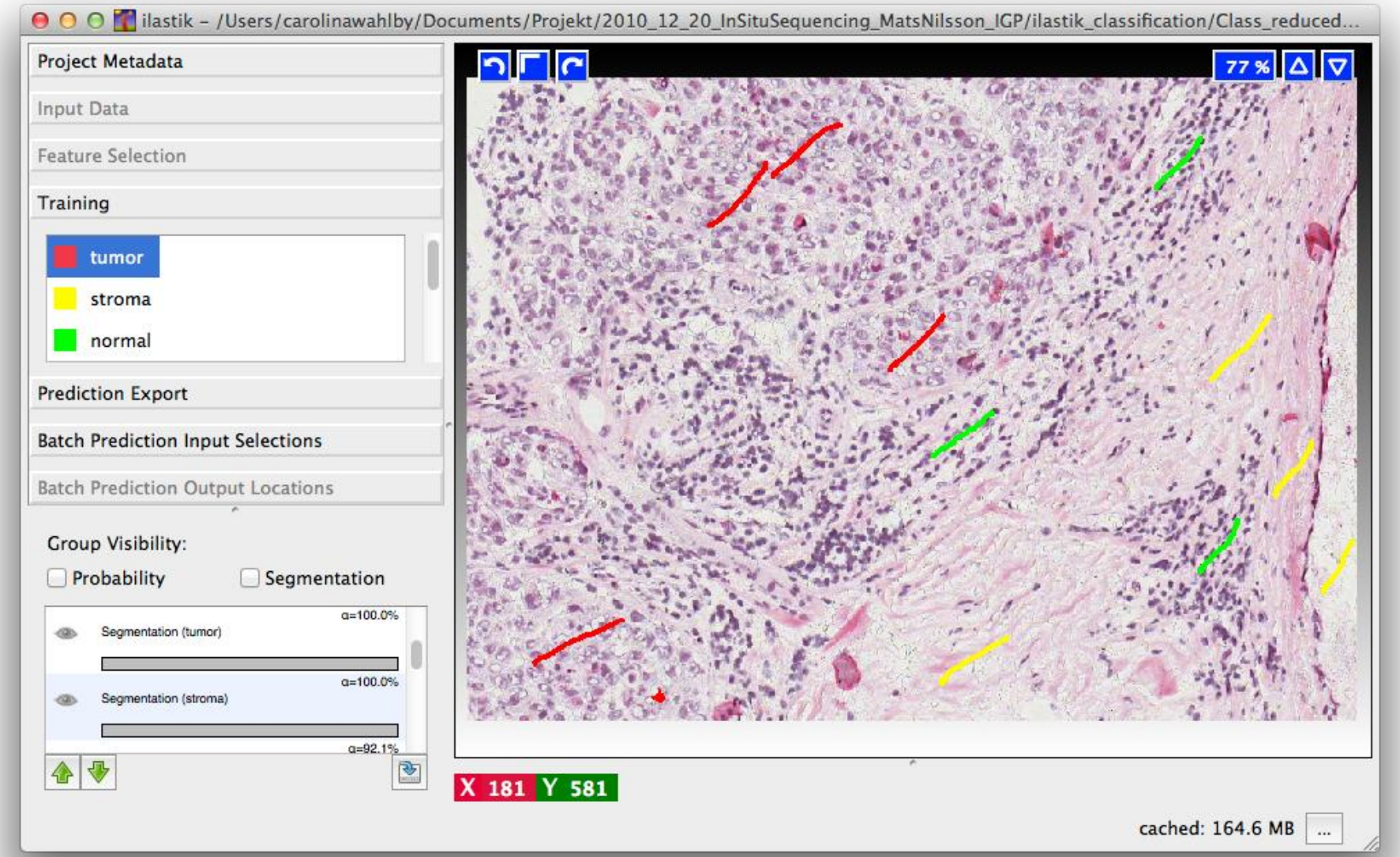
TTGT=HER2 ✱
AGTG=TFRC ✱
GCTA=ACTB +
CAGT=RAC1 o
TTAC=VIM ✱
GTCA=KI-67 ✱
AAGT=STK15 +
GTAA=ER o
... (256 plex)



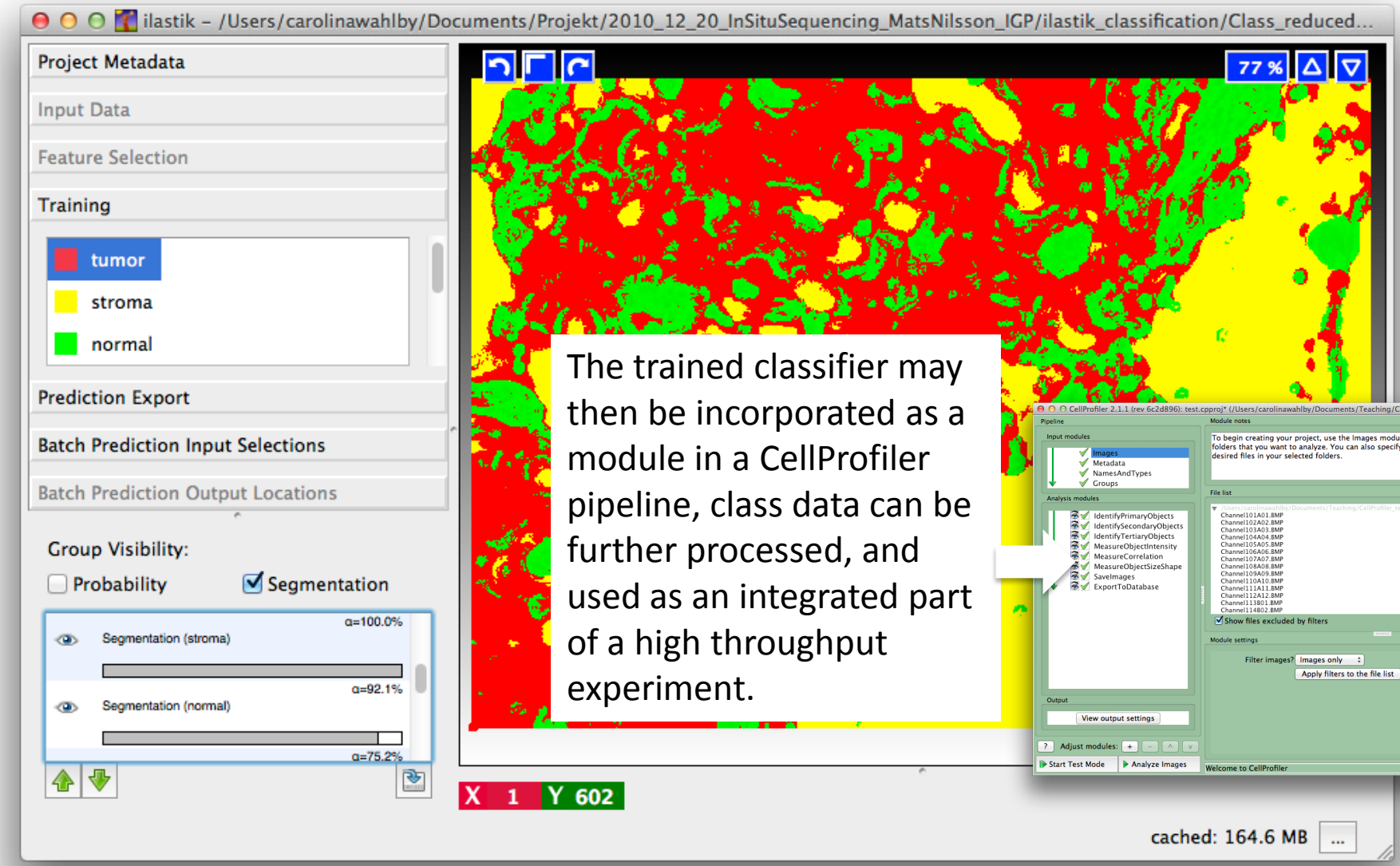


Efficient rigid image alignment was enabled thanks to the RunImageJ module and the TurboReg plugin of ImageJ.

Interoperability: **ilastik**



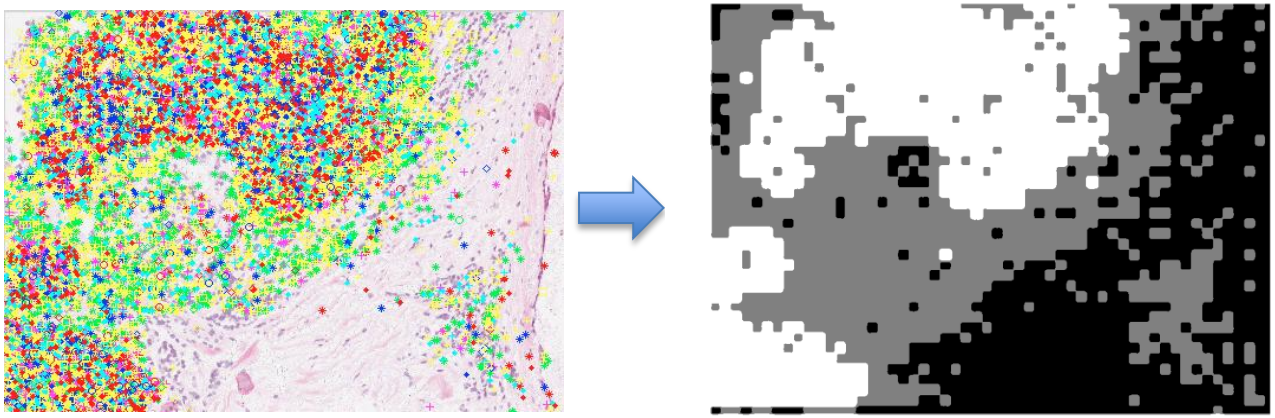
Interoperability: **ilastik**



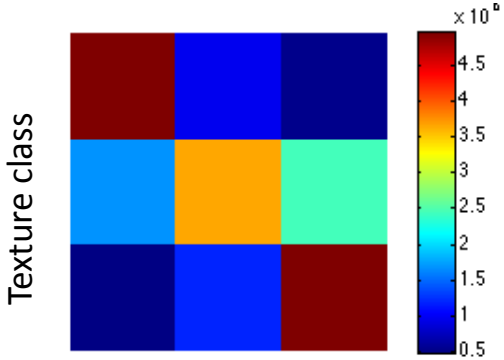
The trained classifier may then be incorporated as a module in a CellProfiler pipeline, class data can be further processed, and used as an integrated part of a high throughput experiment.

Comparing tissue morphology and expression profiling

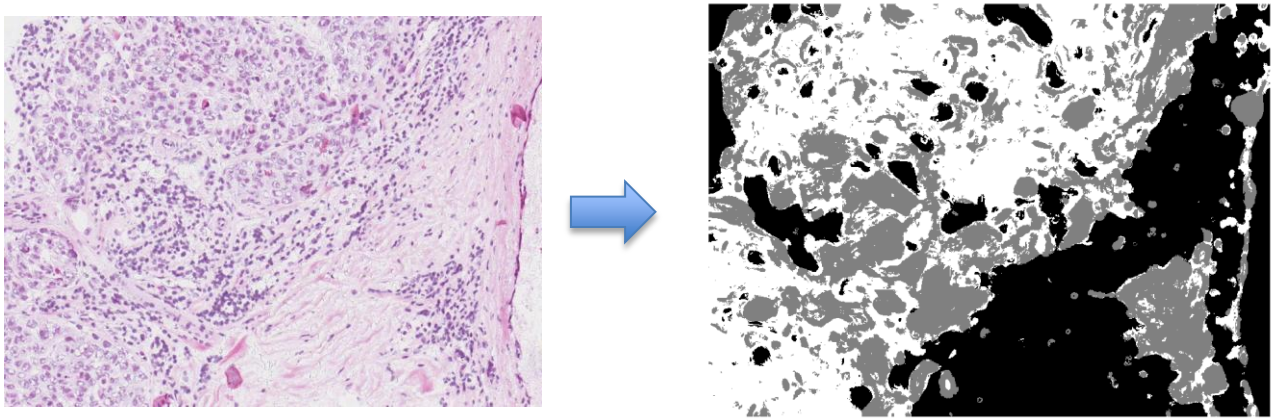
The gene expression map was divided into patches of 100x100 pixels, and transcript counts per bin were normalized by the maximum count of each transcript in any of the patches. Patches were thereafter clustered by k-means clustering (100 random seeds).



Gene expression cluster

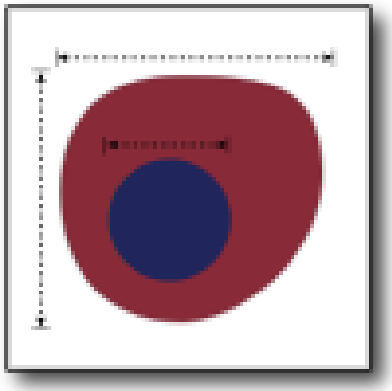


Supervised texture classification (using ilastik)

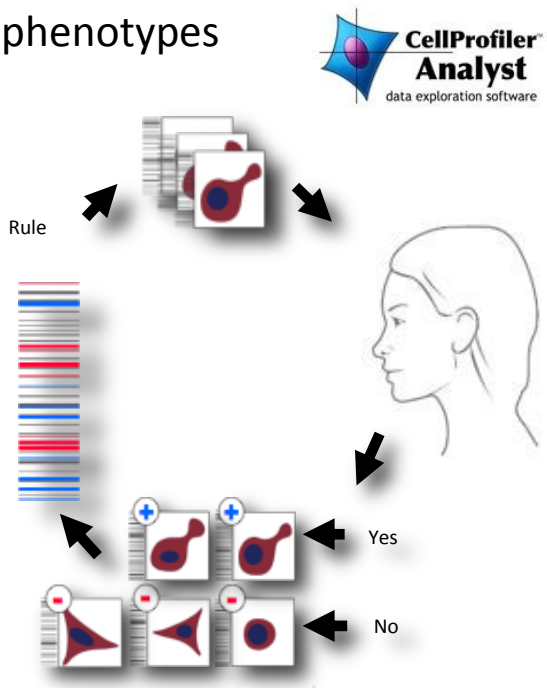


Three approaches for data exploration

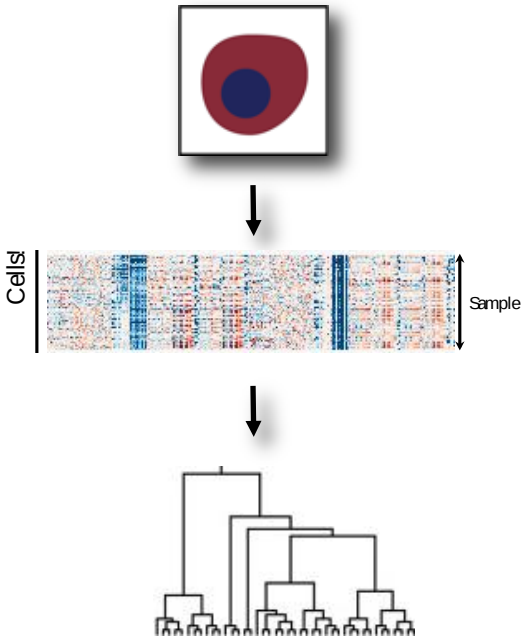
1 Measure known phenotypes



2 Train for known phenotypes



3 Profile to characterize samples

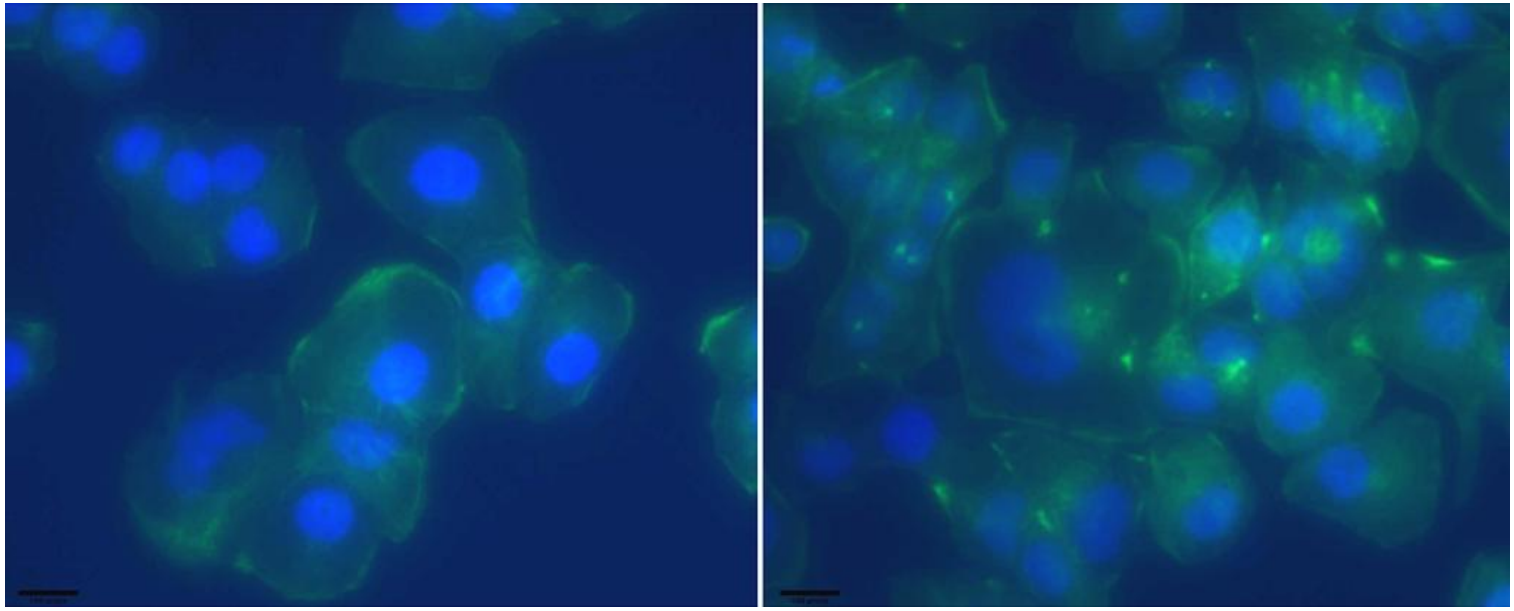


CellProfiler
cell image analysis software

Measure everything, then use machine learning to distinguish a phenotype of interest

Example: Breast cancer

A search for inhibitors of heregulin, which stimulates the ErbB2 pathway and causes cells to be migratory (=bad)

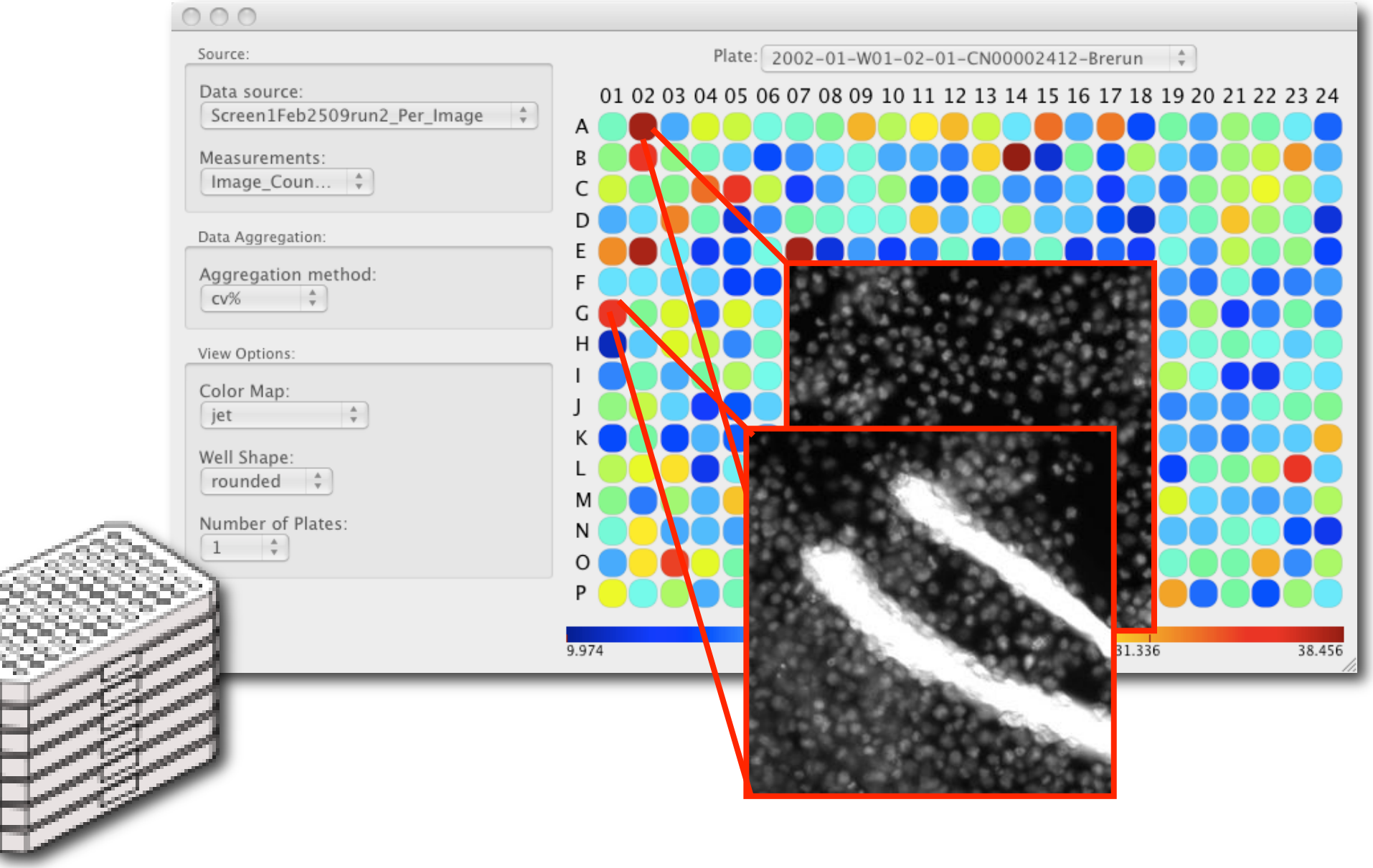


Eric Lander

Piyush Gupta

The phenotypic changes prior to cells becoming migratory are not intuitively described by feature measures; use machine learning to find optimal feature subset!

Quality control: are the hits real?



CellProfiler Analyst for collaborations

For large-scale experiments analyzed on a computer cluster, original data together with analysis output is kept on a server.

Measurements and linked raw data, as well as training samples and overviews of measurements are readily available to all collaborators.

Examples from 2 cell lines
8 plates/cell line
Color=cell count

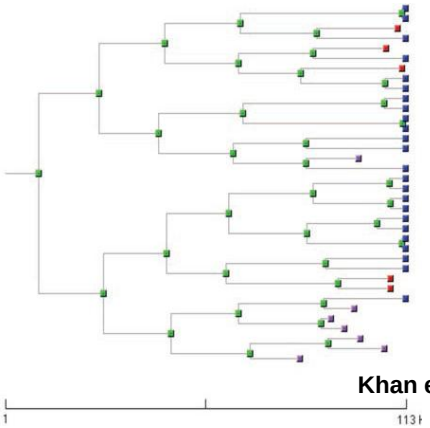
20

Time-Lapse Tool

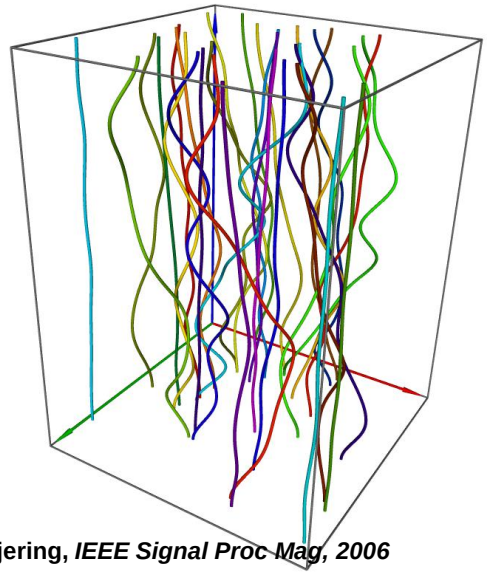
CellProfiler has tracking functionality, based on frame-to-frame object matching or the linear assignment problem (LAP) framework (Jaqaman et al, Nature Methods 2008).

Inspired by other approaches to visualize tracking data, we have recently put together tools for visualizing tracking data.

Lineage tree

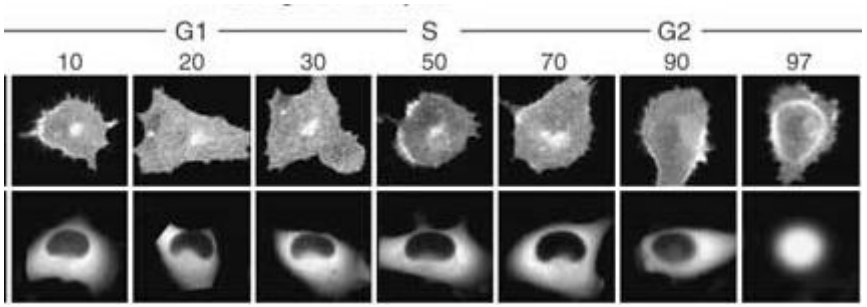


Khan et al, *Cell Cycle, 2007*



Meijering, *IEEE Signal Proc Mag, 2006*

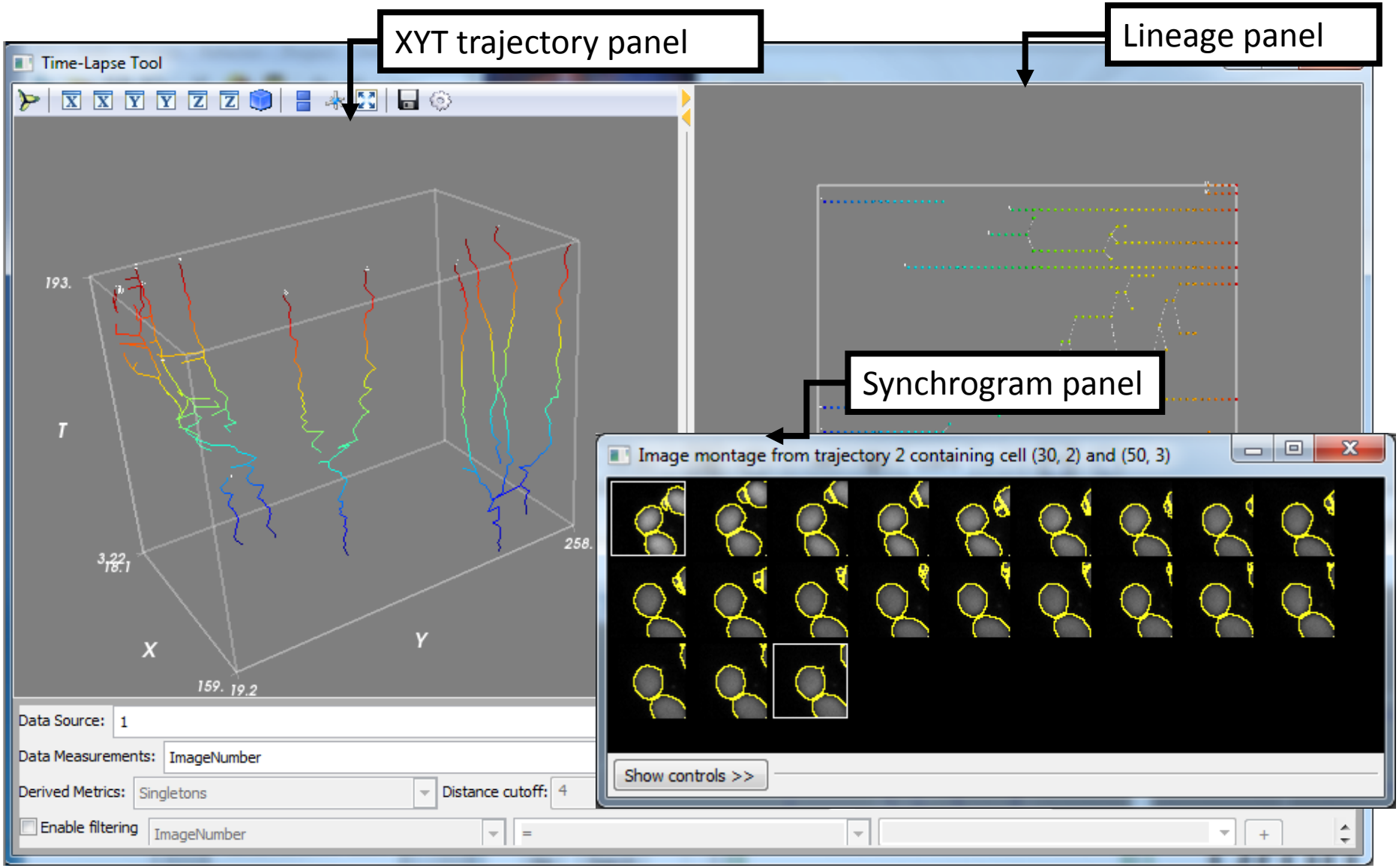
XYT plot



Synchrogram

Sigal et al, *Nat Meth, 2006*

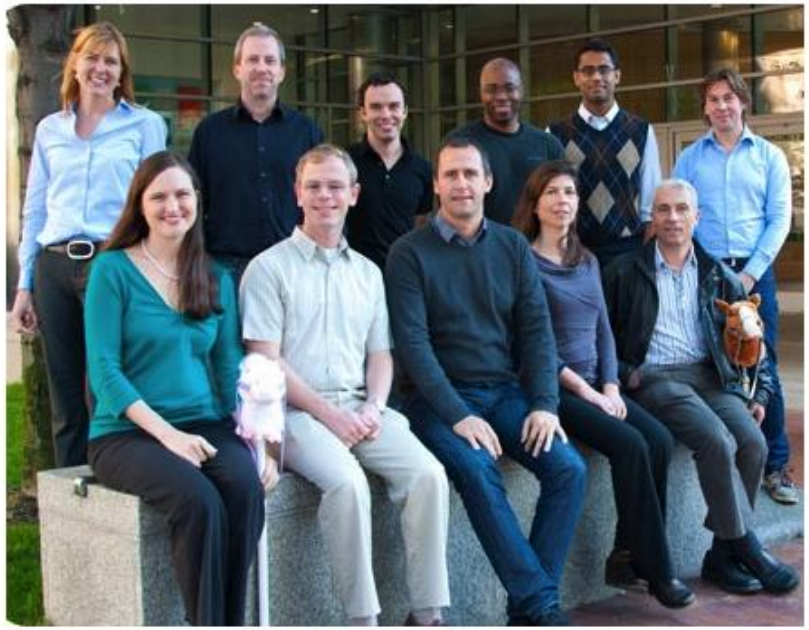
Time-Lapse Visualization Tools



Benefits of open-source software in an interdisciplinary research environment such as SciLifeLab Sweden

- We can assist in seting up an analysis approach on anyone's computer, and quickly help inexperienced users by sharing pipelines and raw data.
- The user community for open-source software is often very responsive and efficient source of help (often better than what can be provided through an expensive service package for a commercial software).
- 'Reproducible research'; we provide our analysis pipelines as part of the supplementary material of published paper.
- Educational value: our computer science students can go in and look at the source, learn, and contribute.

Acknowledgments



Many thanks
to our many
biology collaborators
for images and insights

The CellProfiler project is supported by a grant from the National Institutes of Health (Carpenter: R01 GM089652)

Recent research in the Carpenter laboratory has also been funded by the NSF (Carpenter: CAREER award) and the NIH (Wahlby: R01 GM095672)

BROAD INSTITUTE IMAGING PLATFORM

Anne E. Carpenter

Mark Bray
Auguste Genovesio
Lee Kamentsky

Imtiaz Khan
Vebjørn Ljosa
David Logan

Shantanu Singh
Matthew Veneskey
Carolina Wählby

Contact:
anne@broadinstitute.org

S.D.G.