



HIGH CONTENT SCREENING

CSABA MOLNÁR





HCS - Introduction

- Powerful tool in biological research and drug discovery
- Identifies substances that alter phenotype of a cell
- Simultaneous readout of several parameters
- High throughput thousands of images

Introduction – An example

Motivation

Custumization of healthcare: searching for effective antitumor agents

Goal of screening

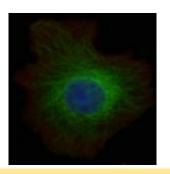
➤ Examination of effects and identification of cancer-relevant cytotoxic compunds

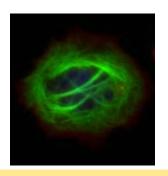
What to measure

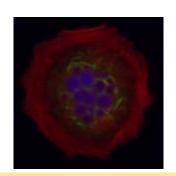
- Observation of phenotipic alterations ->
- > Describe the changes of ,morphological' labels

Dataset

- Cells
 - ➤ MFC-7 breast cancer cells
- Experiment
 - > 133 molecules at 8 concentrations (full dataset)
 - > 2-3 plates with some of these molecules
- Stainings
 - > Cell labeled for:
 - DNA (DAPI), F-actin (Phalloidin) and B-tubulin (anti–β-tubulin antibody)







Workflow/Outline¹

- Image analysis
 - > Illumination correction
 - > Focus detection
 - **➤** Segmentation
 - > Feature extraction
- Quality control field of view, cell-, feature-level
- Preprocessing features
- Annotation
- Downstream analysis

¹Caicedo, Juan C., et al. "Data-analysis strategies for image-based cell profiling." *Nature methods* 14.9 (2017): 849.

Environment

Matlab R2017b

- ✓ Optimized for matrix operations -> perfect choice for image processing
- × Not free..
- ! Free for academic use
- CellProfiler 2.2.0
 - ✓ Free & open source
 - ✓ Easy to build pipelines
 - ✓ Possibility to implement new methods



- ✓ User friendly
- ✓ Machine learning supported (weka core)
- ✓ Helpful functionalities to discover the data



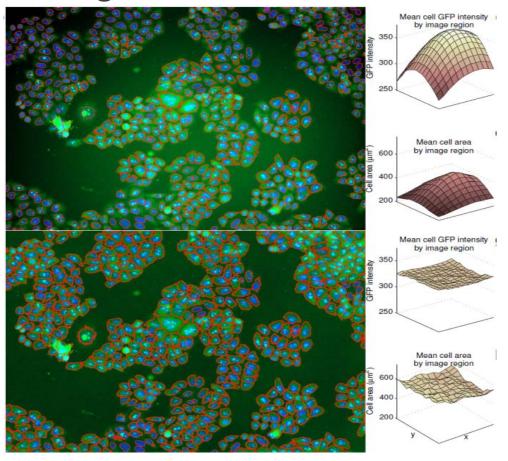


Image analysis I. – Illumination correction

Aim: Remove distortion rising from nonuniform

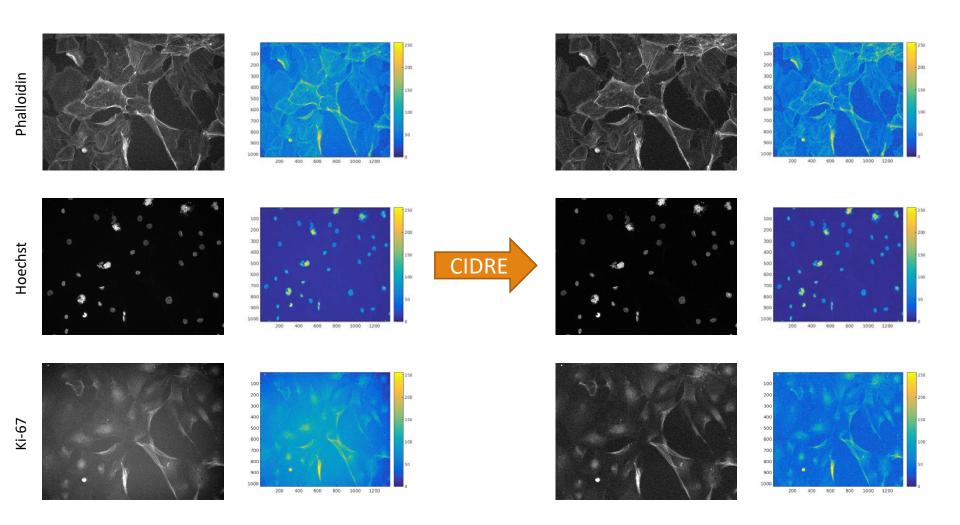
illumination

- Used: CIDRE¹
- Alternatives
 - ➤ Reference image based
 - > Smoothing
 - ➤ Surface fitting
 - ➤ Rolling ball



¹Smith, K., Li, Y., Piccinini, F., Csucs, G., Balazs, C., Bevilacqua, A., & Horvath, P. (2015). CIDRE: an illumination-correction method for optical microscopy. Nature methods, 12(5), 404-406.

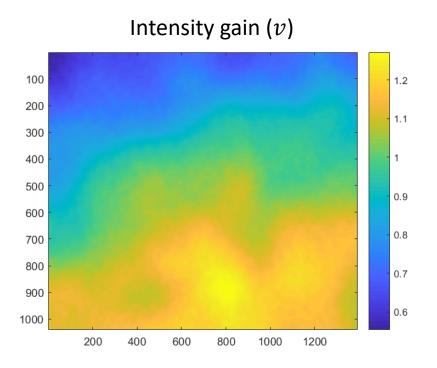
CIDRE for non-uniform illumination correction

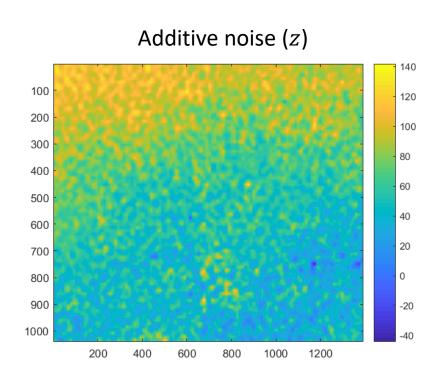


CIDRE for non-uniform illumination correction

$$I_{observed} = (I_{original} - z) * v$$

Real measurements of siRNA screens





Almost 2 factor between the highest and lowest areas!

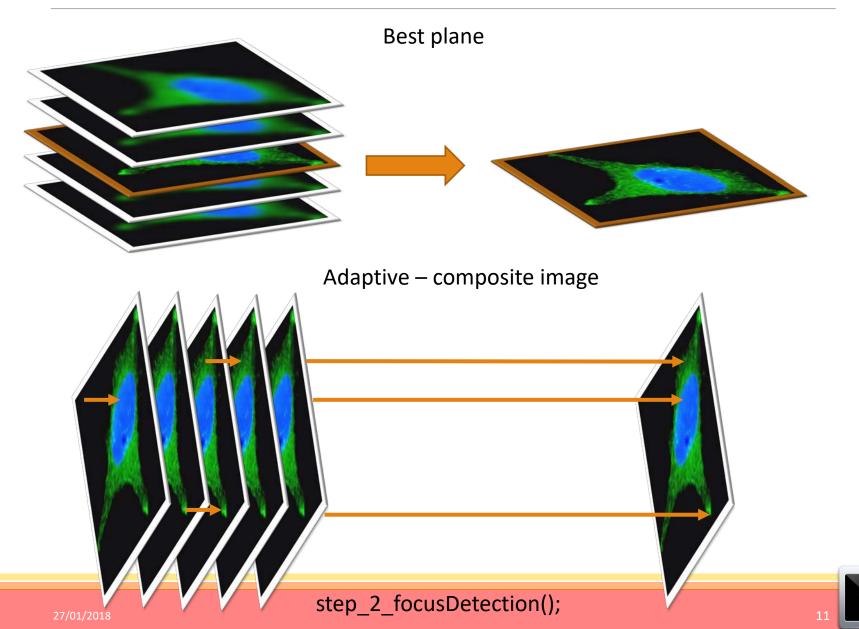


step_1_flatFieldCorrection();

Image analysis II. – Focus detection

- Aim: finding the focused image
 - a. Find the ,most focused' plane
 - Select the image plane that is blurred to the least degree
 - b. Adaptive: create composite image
 - Select the ,most focused' pixel of a position from all planes
- Used: image gradient
 - > Pros: easy to calculate
 - Cons: false selection for high level of noise
- Alternatives
 - ➤ Wavelet transform based (extended depth field)
 - Model based

Image analysis II. – Focus detection

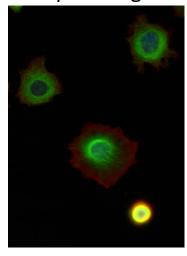


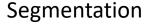
Quality control – Field of view II.

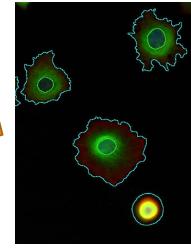
- Discard ,wrong' images
 - Focus score blurred images
 - Used: ratio of image gradient of original and smoothed image
 - > Saturation score oversaturated images
 - Number of saturated pixels

Aim: Extraction of regions (set of pixels) belonging to the same object
Input image

- Steps
 - 1. Nuclei segmentation
 - Used: model based approach
 - Thresholding, edge detection,
 - 2. Cell segmentation
 - Based on nuclei as seeds
 - Used: adaptive thresholding
 - Region growing, simple extension etc.
- Alternatives
 - Machine learning based (Ilastik)
 - Manual labeling of pixels
 - Training a classifier







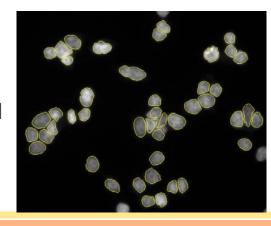


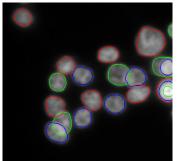
27/01/2018 1₂

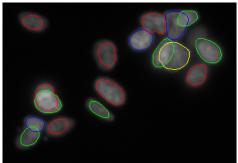
- Results from our group: variational shape models
 - ➤ Multi-layered ,gas of circles'
 - For overlapping, near-circular objects

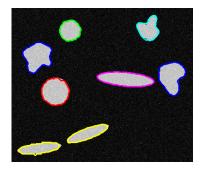


- Computationally cheap shape model for circular, elliptical or amoeba shaped objects
- Splitting active contour
 - Higher-ordered active contour for declumping





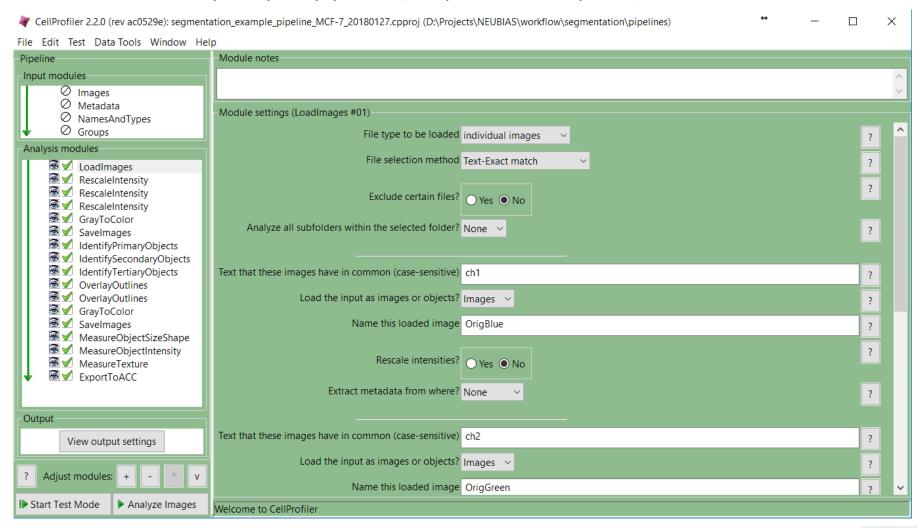




- Alternatives: template based methods
 - Hough transform
 - Ovusculae



Run CellProfiler. Compose your pipeline (or open the example ©).





Run old CellProfiler. Compose your pipeline (or open the example ©).

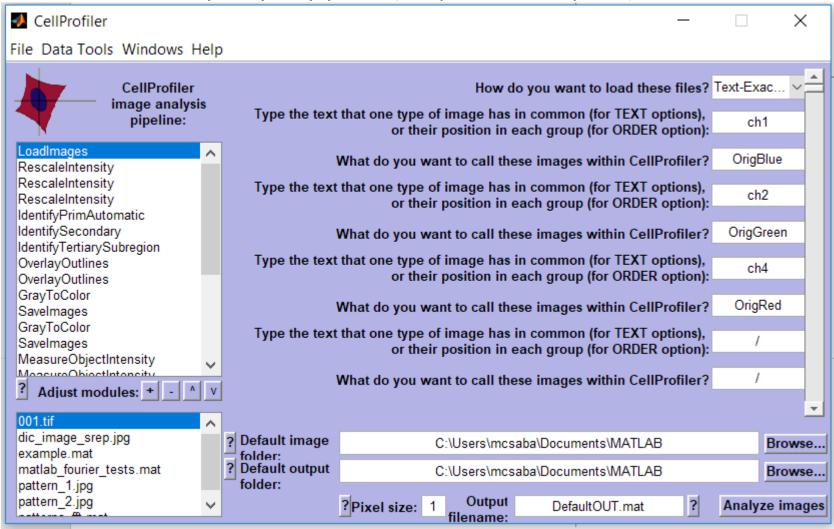
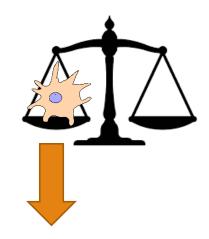


Image analysis IV. – Feature extraction

- ,Measure' objects
 - > Shape features
 - Description of object boundary
 - Area, perimeter, roundness etc.
 - Used: Matlab/Python regionprops+Zernike moments
 - > Intensity features
 - Pixel-level statistics on image intensities
 - Mean, min, max etc.
 - Used: Matlab/Python basic statistics
 - > Texture features
 - Spacial arrangement of intensities: patterns
 - Contrast, correlation, entropy etc.
 - Used: Haralick and Gabor features
 - Context' features*
 - Information about environment
 - Number of neighbours, mean distance of neighbours etc.



features

| 3.1120650e+01···7.7773434e+02···3.1434293e+02···7.8409 |
|--|
| 1.6845960e-02···1.9971008e-01···5.5214560e+01···5.4987 |
| 6.2421110e-04···5.3711757e-03···1.0009918e-02···6.1241 |
| 2.7683423e-02···1.1028604e-02···2.0019837e-02···8.3985 |
| 3.5639500e+02···3.9441096e-01···9.4748706e+01···5.6942 |
| 4.2421598e+00···1.5556303e+01···1.5457754e+00···1.8102 |
| 1.1998827e-01···9.9427325e-01···7.0792703e-01···1.6776 |
| 9.5981506e-01··-3.2124113e-01···7.9436919e-01···1.7333 |
| 4.7491137e+00···4.6793522e+00···2.0644165e+00···2.8551 |
| 2.4482429e+02···7.4618803e-02···4.5356899e-02···1.6845 |
| 4.4598201e+01···2.9085893e+01···8.8649476e-03···2.4793 |
| 2.3437858e-02···5.0604858e+01···1.3568347e+02···4.1354 |
| |

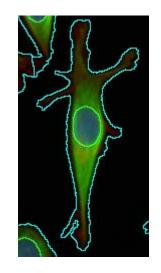


Quality control II.

- Field of view QC
 - ➤ Cell number discard empty images

- Cell-level QC
 - Discard outlier objects/cells (e.g. wrong segmentations)
 - > Rare phenotypes can stay unrevealed
 - ➤ Better included in further steps

Rare phenotype



Segmentation error





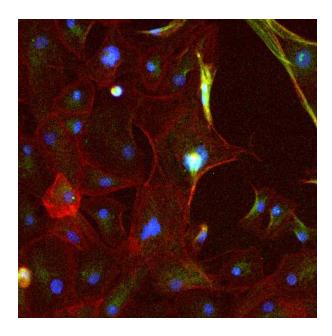
Preprocessing features

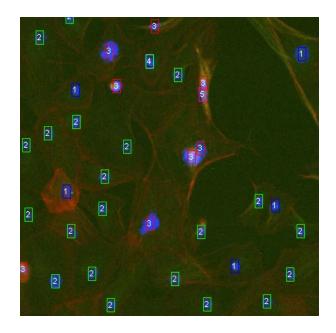
- Dimensionality reduction
 - Aim: remove redundant features, keep most informative ones
 - ➤ Used: Ranking features by their contribution to reduce the entropy (InfoGainAttributeEval from weka)
 - > Alternatives:
 - Principal Component Analysis
 - Factor analysis
 - Reduce by pairwise similarity



Discover data

- Aim: Identify subgroups (,phenotypes') Used: supervised machine learning
- Alternatives:
 - > Identify by clustering unsupervised machine learning



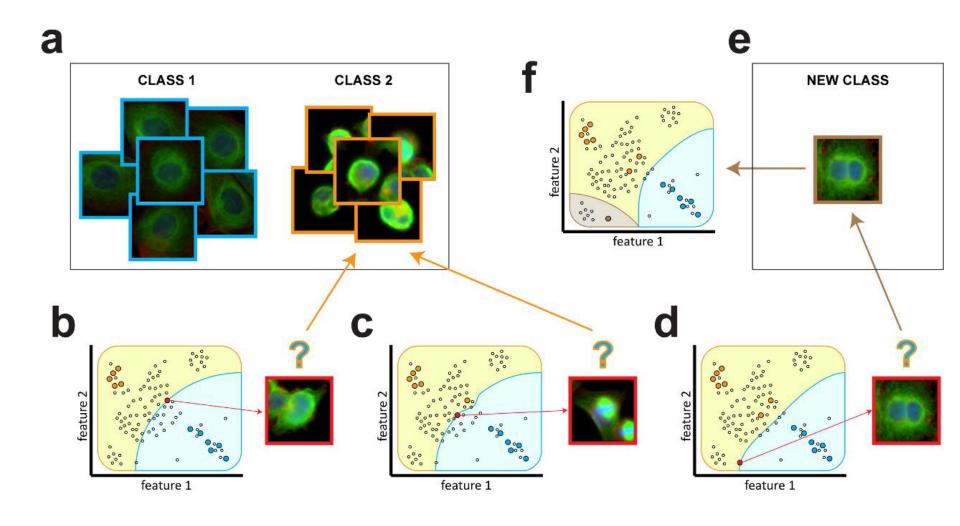


Phenotyping tool - ACC

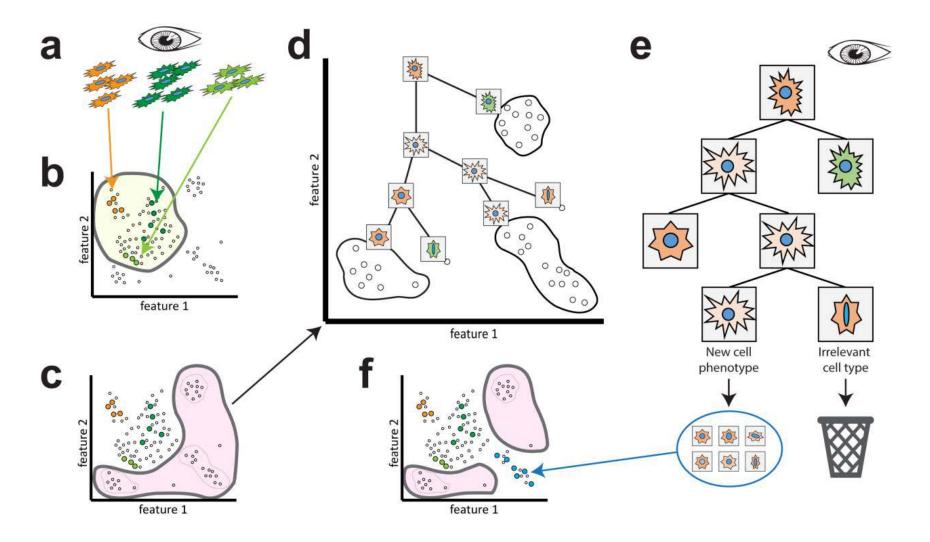
- Examination of the feature space
- ADVANCED CELL CLASSIFIER

- Classical annotation
 - Known phenotypes
 - ➤ Labeling cells
- Active learning
 - ➤ Smart questions about the uncertain decisions
- Discovery of new phenotypes
 - > Reveal the outliers
- Find similar cells

ACC – Active learning

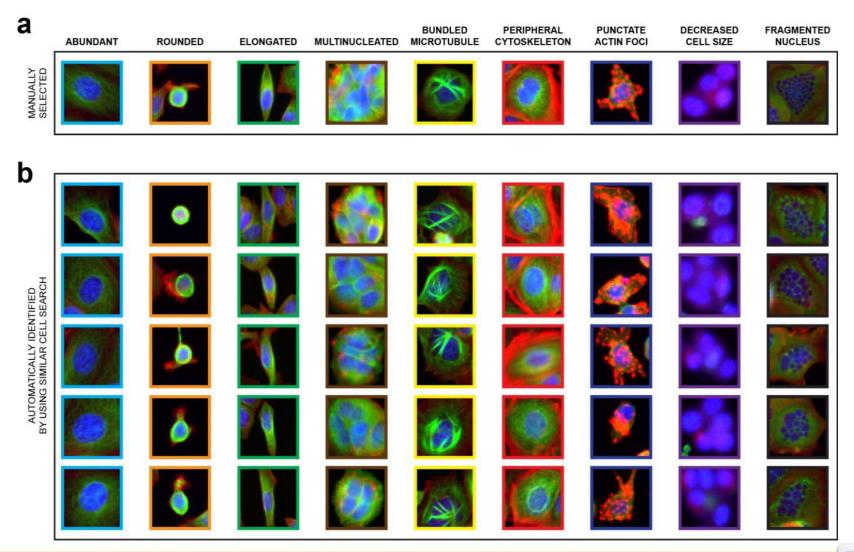


ACC - Phenotype finder



23

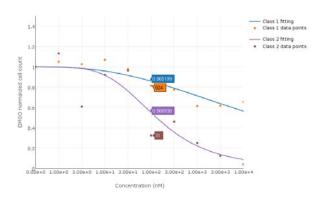
ACC - Find similar cells

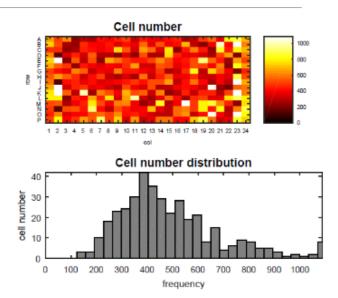


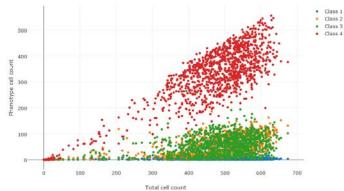
>_

Downstream analysis

- Mean/median profiles
- Classifications
- Feature based statistics
- Comparison to controls
- Similarities of profiles
- Visualisation







>_