Phenotypic high-content screening utilizing multi-parametric data analysis for novel lead identification

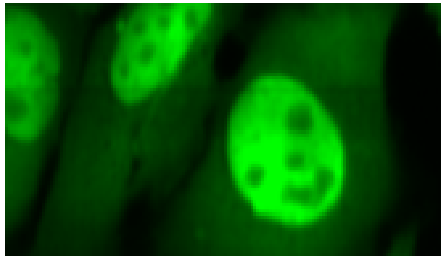
Daniela Gabriel, Anne Kümmel, Christian Parker

January 13th 2010

NOVARTIS

Phenotypic high-content screening

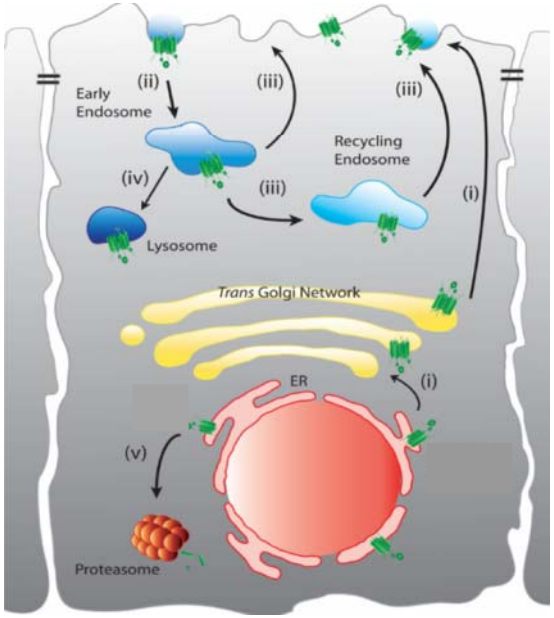
- In HCS, generally cells are analyzed with regard to target specificity
- A great potential of HCS lies within the analysis of cellular phenotypes by generation of multidimensional readouts of cellular effects in response to compound treatment
- Multivariate statistics provide a range of data reduction and classification tools to not only identify hits but also to classify the compounds effect and to consider different responses in sub-populations
- Utilizing multivariate analysis of phenotypic profiles enhances the potential of hit discovery in small molecule screening and help classifying hits for target identification



Introduction

Primary screening data

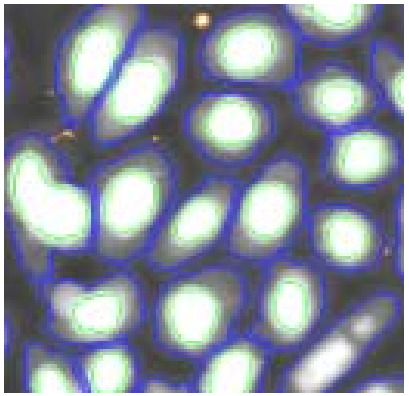
- A Cl^- selective ion channel is expressed in epithelia cells
- Point mutation is the most common form of the disease
- Due to mutation the folding in ER is improper and protein is rapidly degraded
- Aim of the assay was to find correctors to facilitate the incorporation of the mutated channel into the membrane
- High-content primary screen of 100k cpds



Introduction

Experimental setup and image analysis

- Cellomics image analysis (BioApplication Morphology.V3)
  - Cytoplasm (Ch1): Draq5 „background“ staining
  - Nucleus (Ch2): Draq5 staining
  - Target (Ch3): antibody staining in cell area (nucleus + cytoplasm)
- 35 parameters determined
  - 11 cell shape measures
  - 9 nuclear shape and intensity measures
  - 15 fluorescence intensity staining measures



- 600 cells per well analyzed
- Test set of 31 plates containing ca. 10 k out of 100 k samples (~10%)

Fluorescent vs multiple parameters

InCell3000 vs. Cellomics analysis

INCell 3000 analysis based on fluorescent parameters only

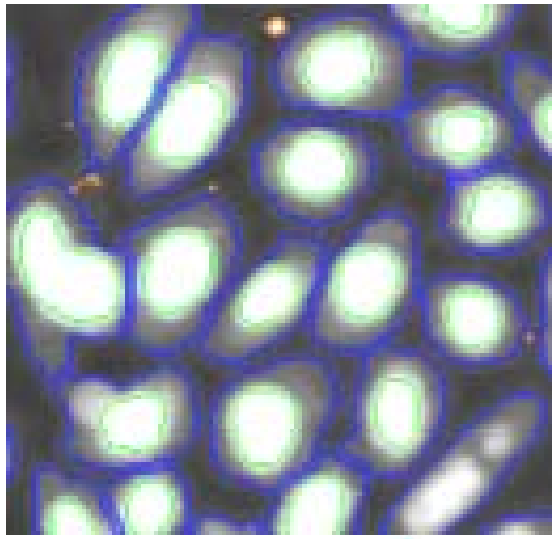
- I pos rings: hits > 20% activity corrected
- N pass: toxic cpds <70% of DMSO Npass
- I pos nucleus: fluorescent cpds

➔3 parameters in total

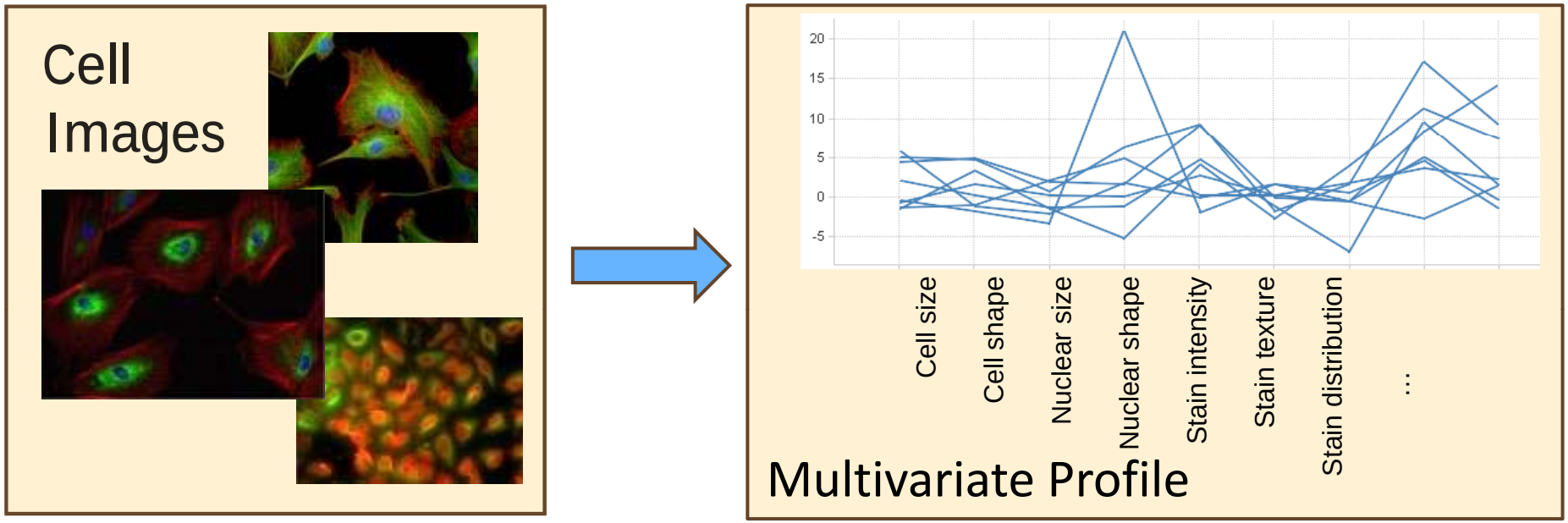
Cellomics analysis based on multiple parameters

- 11 cell shape measures
- 9 nuclear shape and intensity measures
- 15 fluorescence intensity measures

➔35 parameters in total



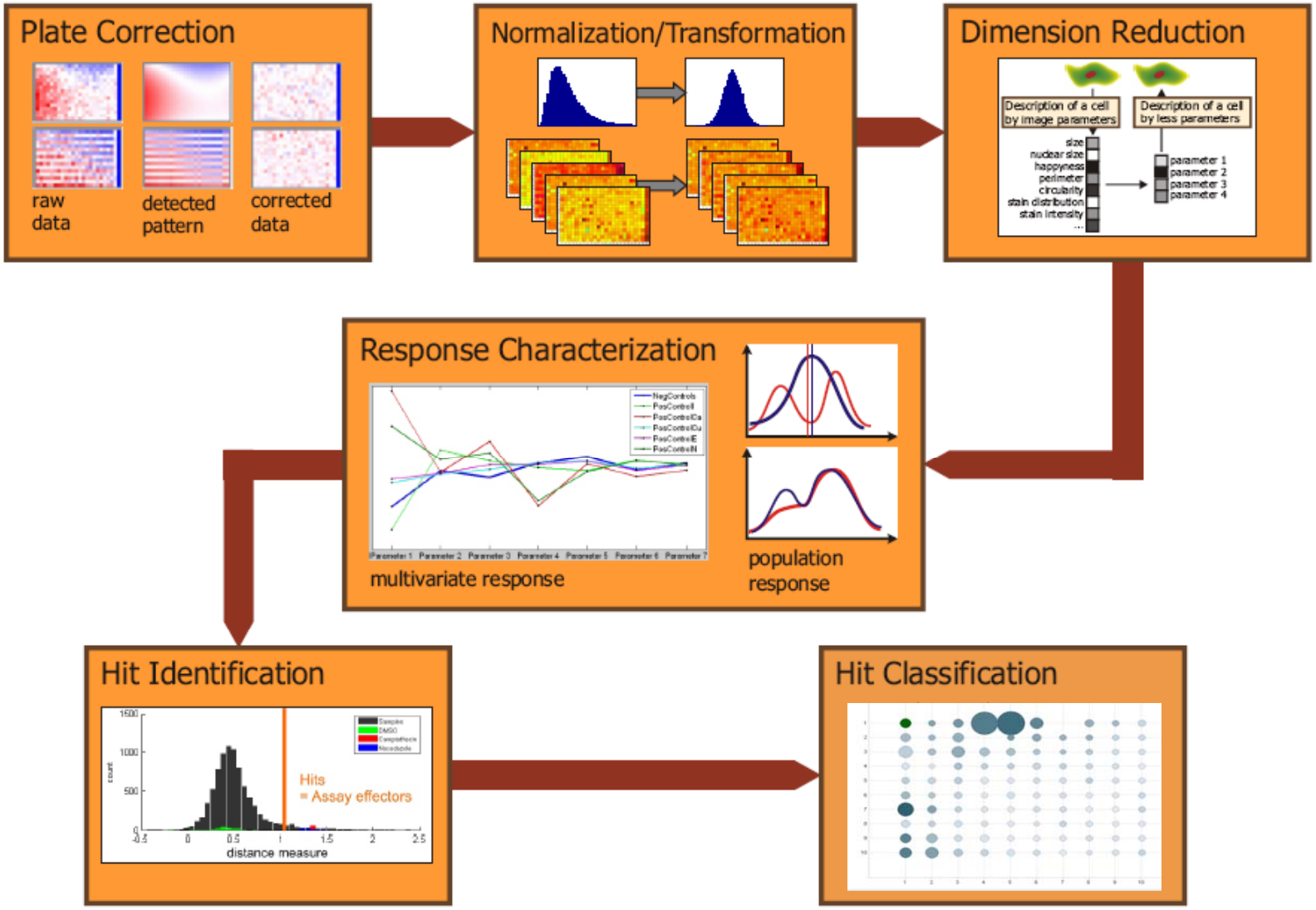
Multivariate profile analysis



- Tools to process and access the multivariate profiles
- Evaluation of methods suitable to explore multivariate profiles and select hits/leads based on multiple readouts

Multi-parametric data handling

Overview



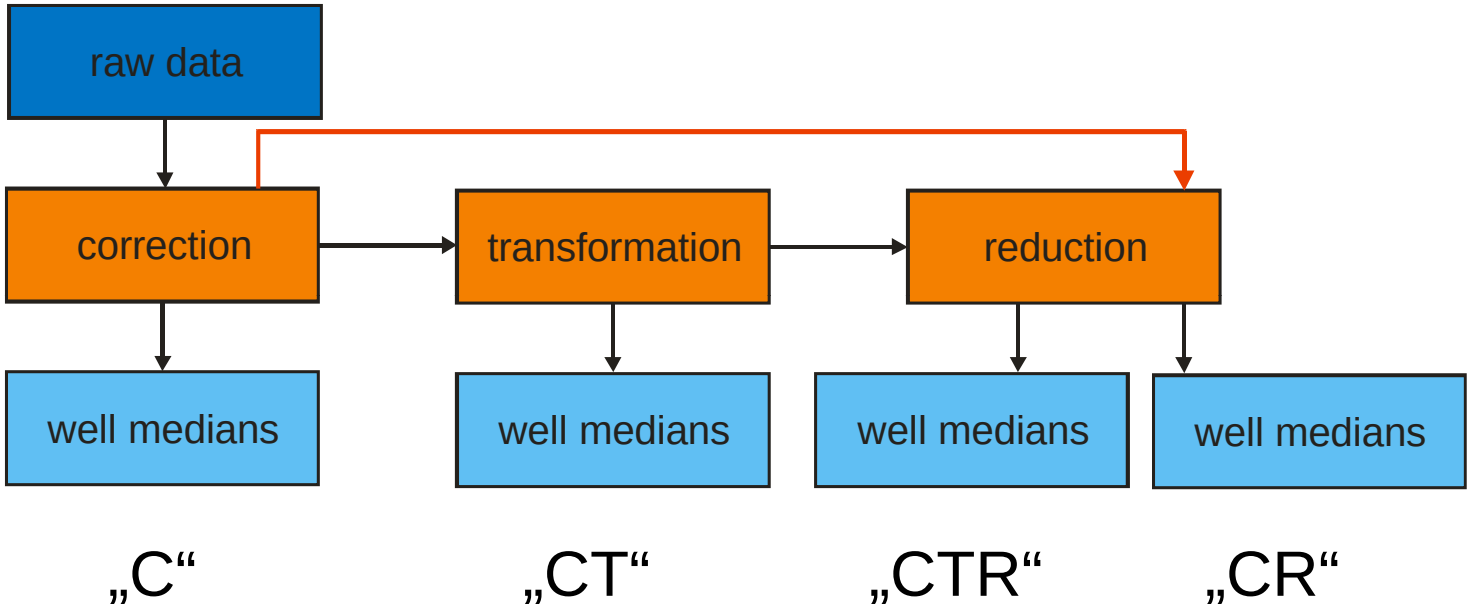
Dimension reduction

Factor analysis, Principal component analysis (PCA)

- Factor analysis is applied to map features into a reduced dimensional space by a set of factors that reflect the major phenotypic attributes
- Factor analysis seeks to account for the common variance, which is regarded as that variance shared among variables
- PCA seeks to reduce the dimensionality into a small number of dimensions that maximally accounts for the total variance
- Both in Factor analysis and PCA the components are modeled as linear combinations of either the latent underlying factors or the measured variables, respectively

Data pre-processing

Different pre-processing setups



- Is information lost when reducing the dimensions?
- Is transformation beneficial for subsequently applied methods that assume Normal distribution?

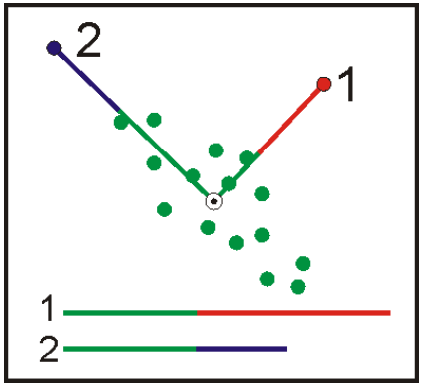
Overview of analysis approach

1. Plate correction, Normalization to DMSO control
2. Well median calculation
3. Hit pre-selection: Selection of compounds based on Mahalanobis distance
4. SOM Clustering: Based on their multivariate effect in unsupervised manner using self-organizing maps (SOMs)
5. Profile exploration of clusters: Based on profiles and images to determine „interesting“ groups

Hit pre-selection

Selecting compounds dissimilar to negative controls

- Mahalanobis distance: Non-specific measure on dissimilarity of a sample to a group of samples



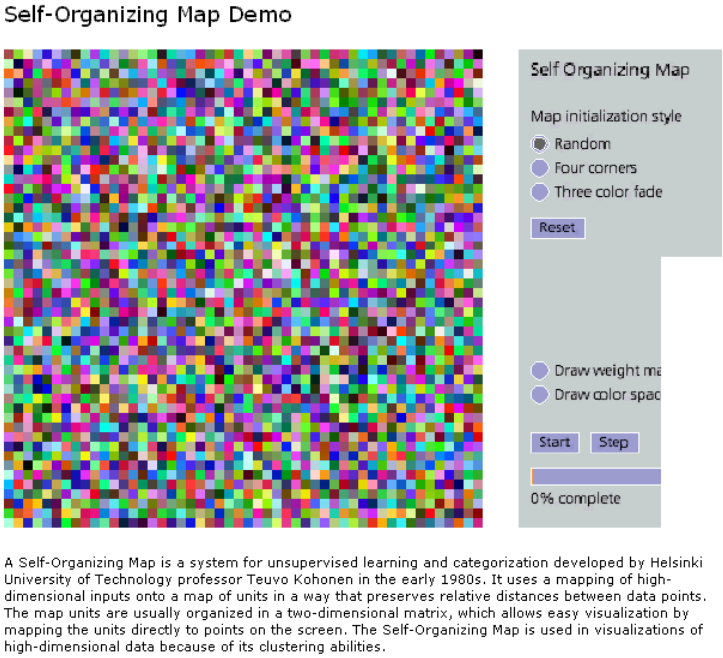
Green dots: DMSO
Red dot (1): compound 1
Blue dot (2): compound 2

- The scaling was done to account for variations: compound 2 has a shorter Mahalonobis distance due to the higher variation within the DMSO in that direction
- Threshold used in following example: upper 5% resulting in 521 compounds

„Hit“ exploration with unsupervised clustering

Self-organizing maps

- Grid of nodes having certain attributes
- Allocate samples to the nodes according to the attributes
- Machine Learning: Determine the node attributes such that
  - there are nodes to which the samples can be allocated to, and
  - neighboring nodes are similar to each other

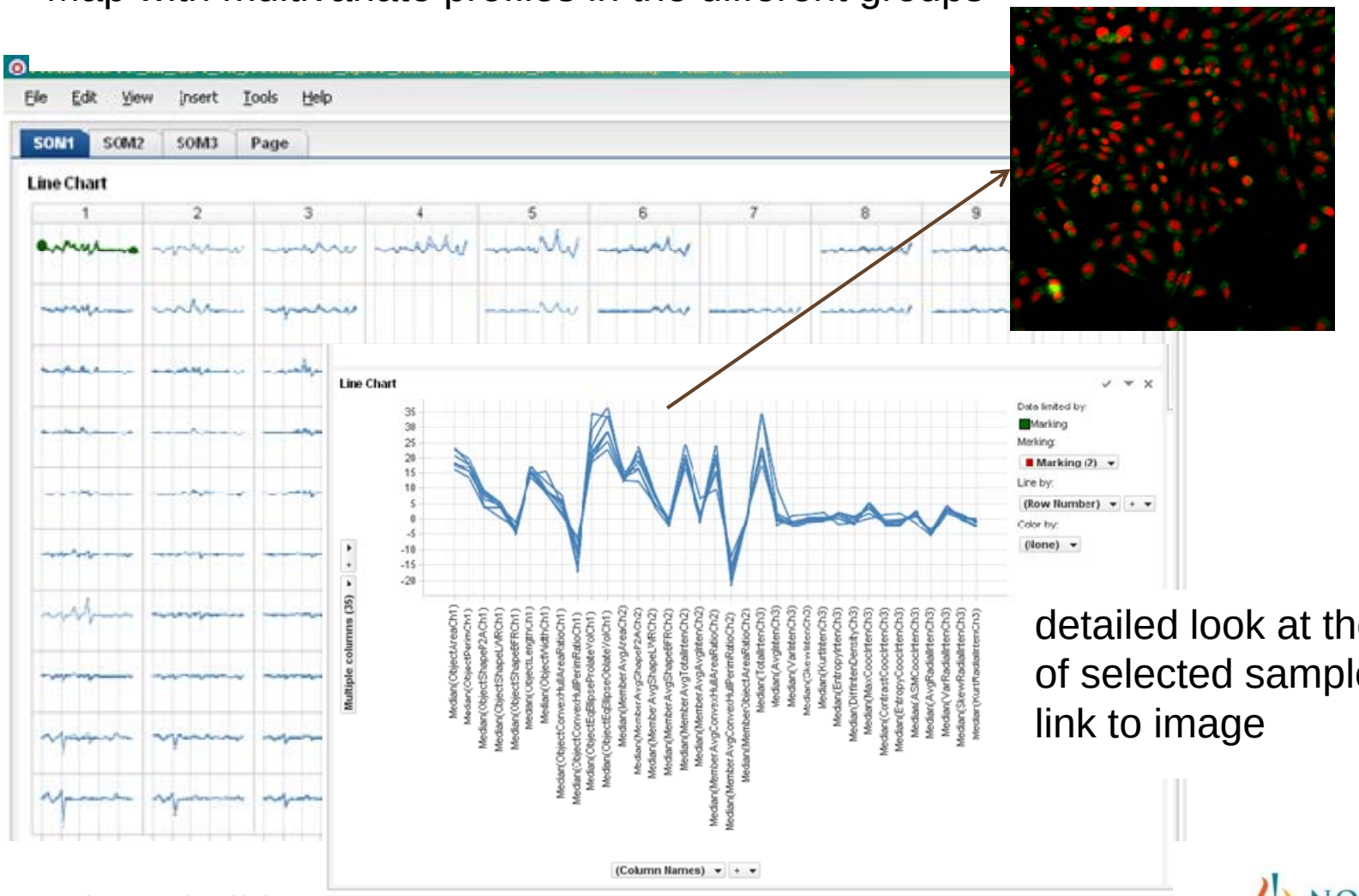


Demo taken from http://www.thbbpt.net/sketches/som/

NOVARTIS

Visualization for hit group analysis

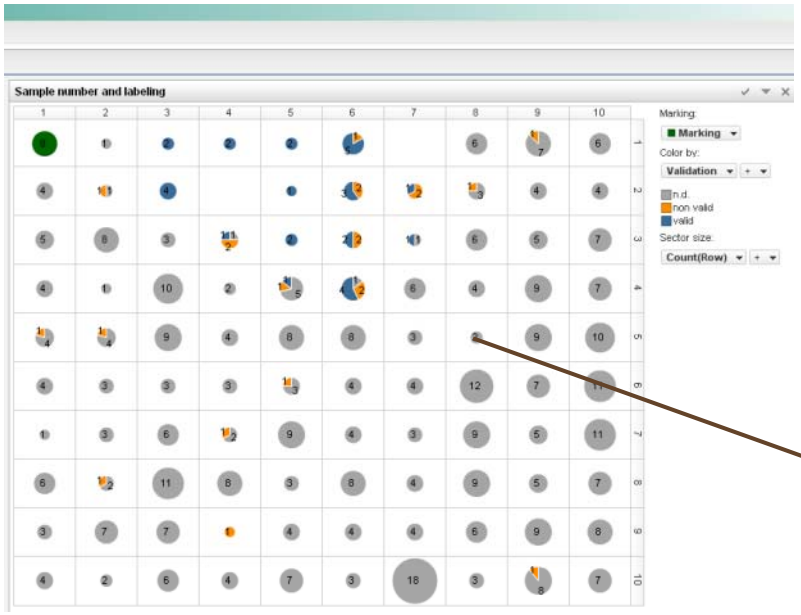
Map with multivariate profiles in the different groups



detailed look at the profile of selected samples with link to image

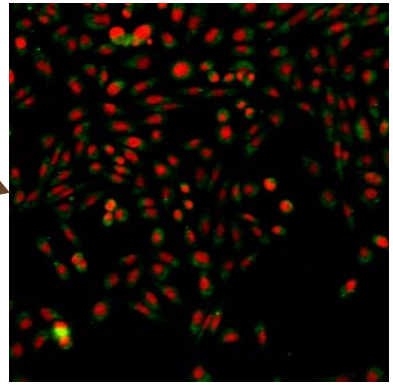
NOVARTIS

Visualization for hit group analysis

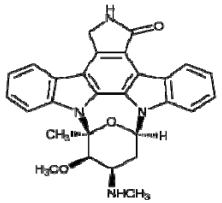


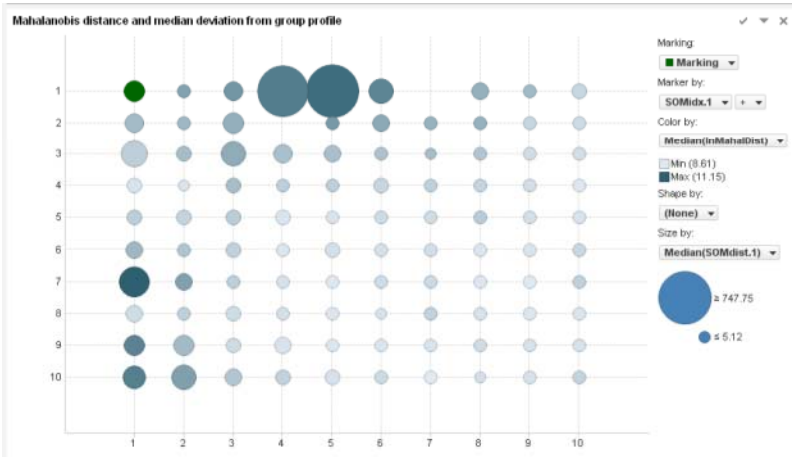
Pies indicating number of samples and sample labeling

Blue: confirmed in univariate analysis
Yellow: not confirmed in univariate analysis



With link to image and compound structure





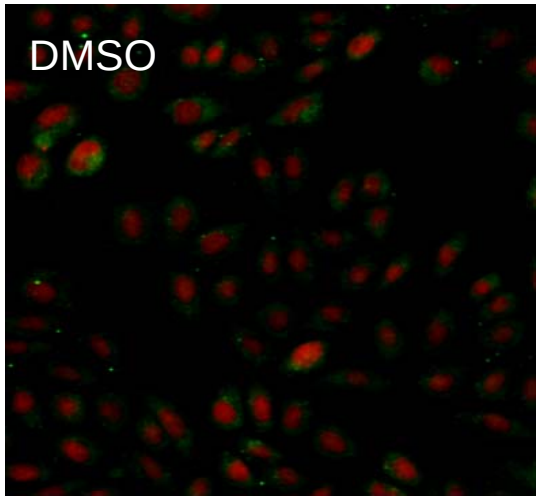
Pies indicating the Mahalanobis distance and average dissimilarity of group members to group prototype

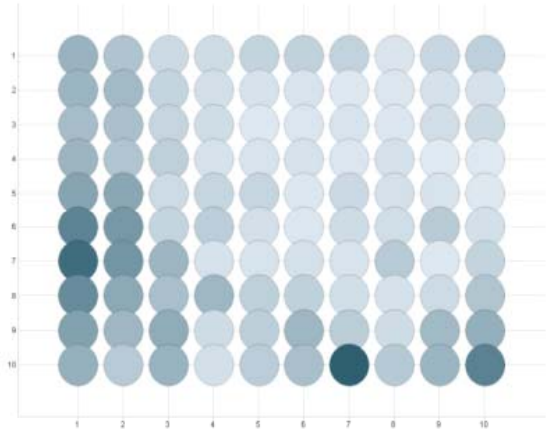
NOVARTIS

„Hit“ exploration based on SOM

DMSO control image

- Exploration of groups that have remarkable Mahalanobis distance: the darker the more distant to DMSO



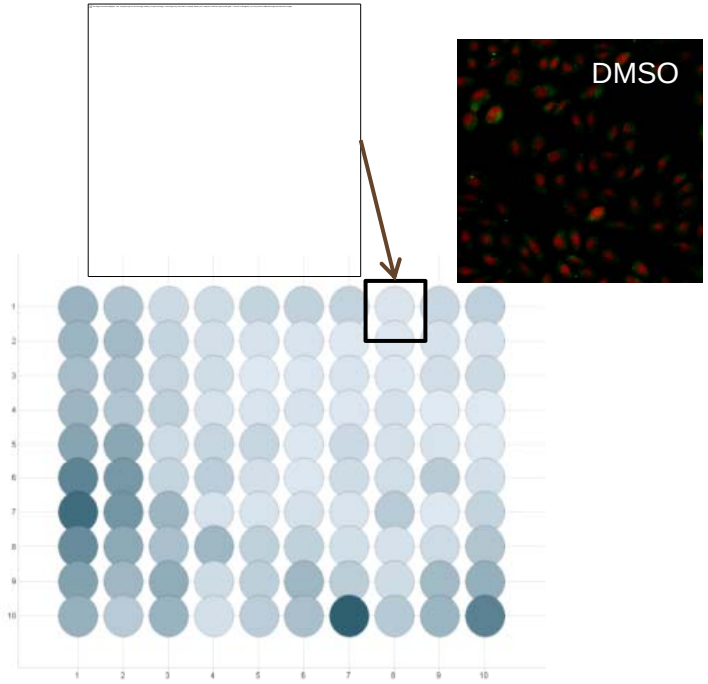


„Hit“ exploration based on SOM

Low Mahalanobis distance

- Exploration of groups that have remarkable Mahalanobis distance: the darker the more distant to DMSO

- Low Mahalanobis distance

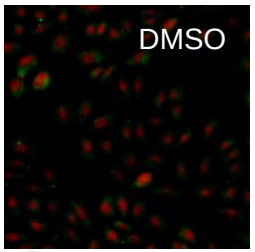


NOVARTIS

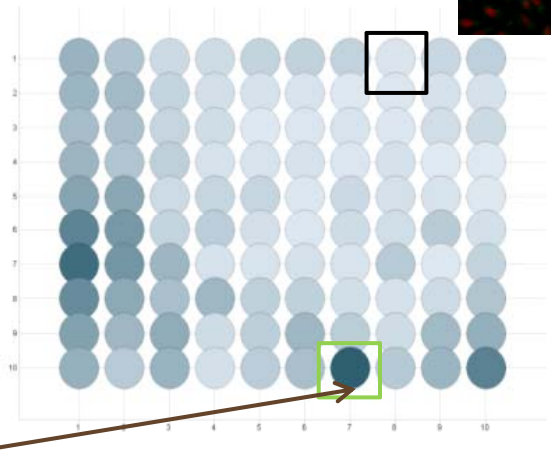
„Hit“ exploration based on SOM

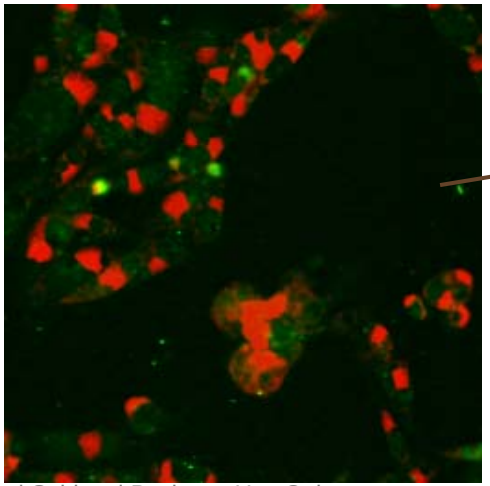
Single high Mahalanobis distance

- Exploration of groups that have remarkable Mahalanobis distance: the darker the more distant to DMSO



- Low Mahalanobis distance
- Single high Mahalonobis distance

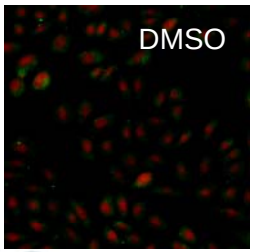




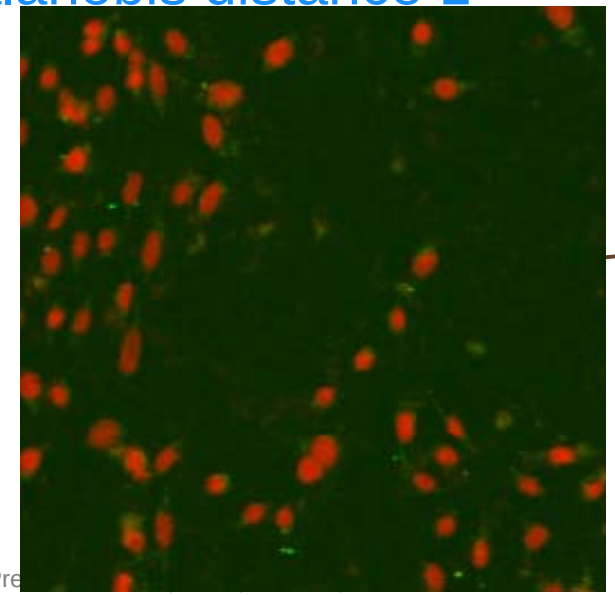
„Hit“ exploration based on SOM

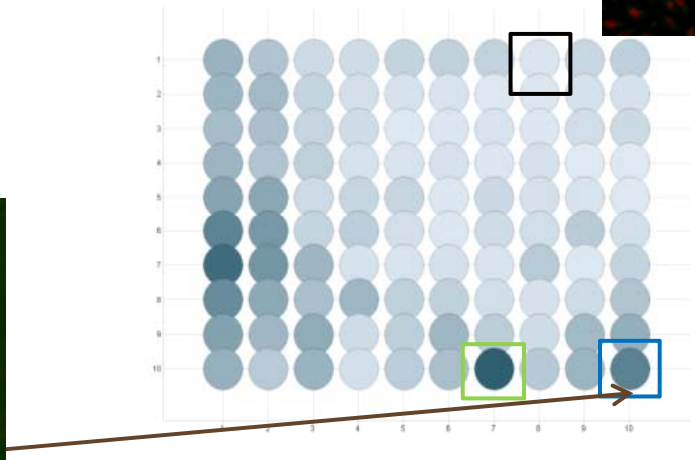
High Mahalanobis distance 1

- Exploration of groups that have remarkable Mahalanobis distance: the darker the more distant to DMSO



- Low Mahalanobis distance
- Single high Mahalonobis distance
- High Mahalanobis distance 1



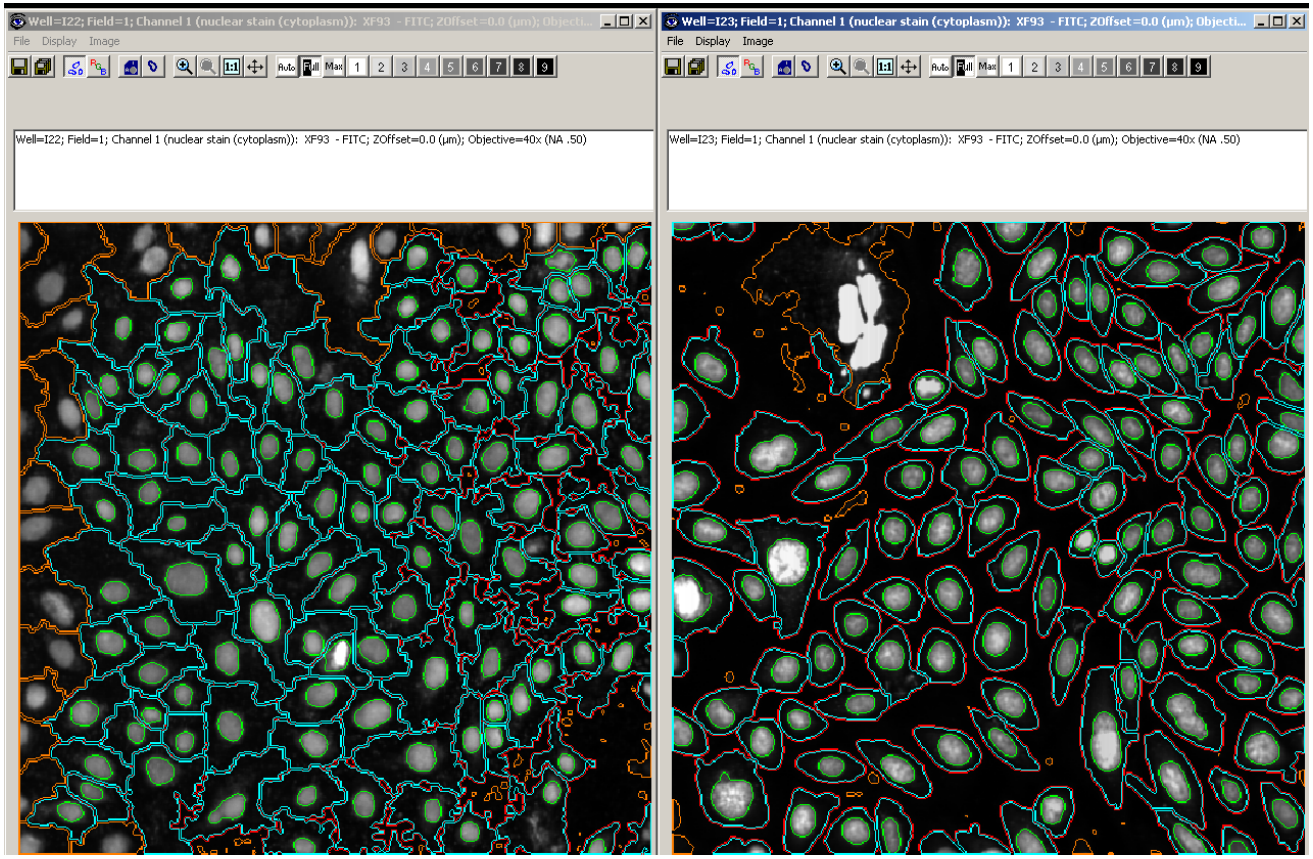


cells with protrusions

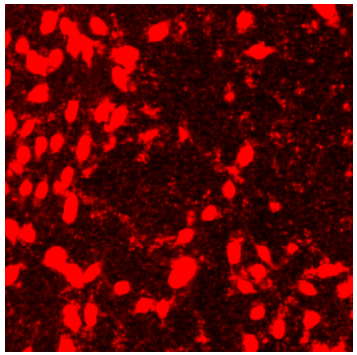
NOVARTIS

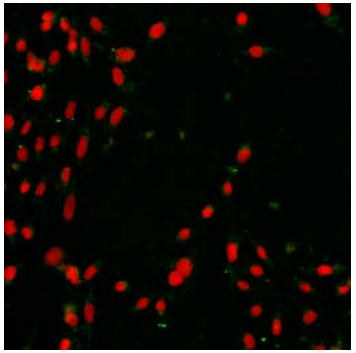
„Hit“ exploration based on SOM

High Mahalanobis distance 1: bacteria disturb the image analysis



Amplified Draq5 stain



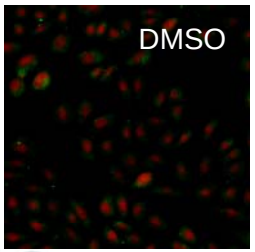


NOVARTIS

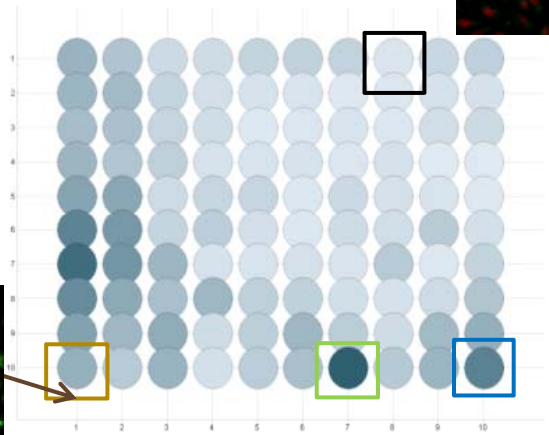
„Hit“ exploration based on SOM

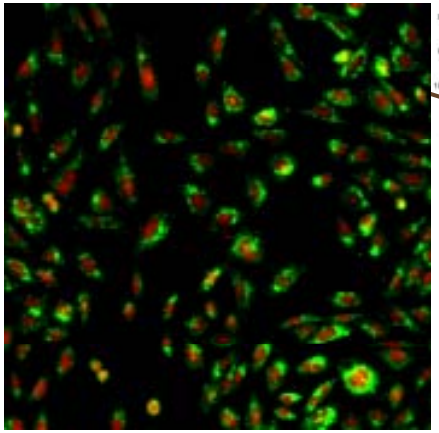
Group of 8 compounds

- Exploration of groups that have remarkable Mahalanobis distance: the darker the more distant to DMSO



- Low Mahalanobis distance
- Single high Mahalonobis distance
- High Mahalanobis distance 1
- Group of 8 compounds





Large cells with non-convex nuclei

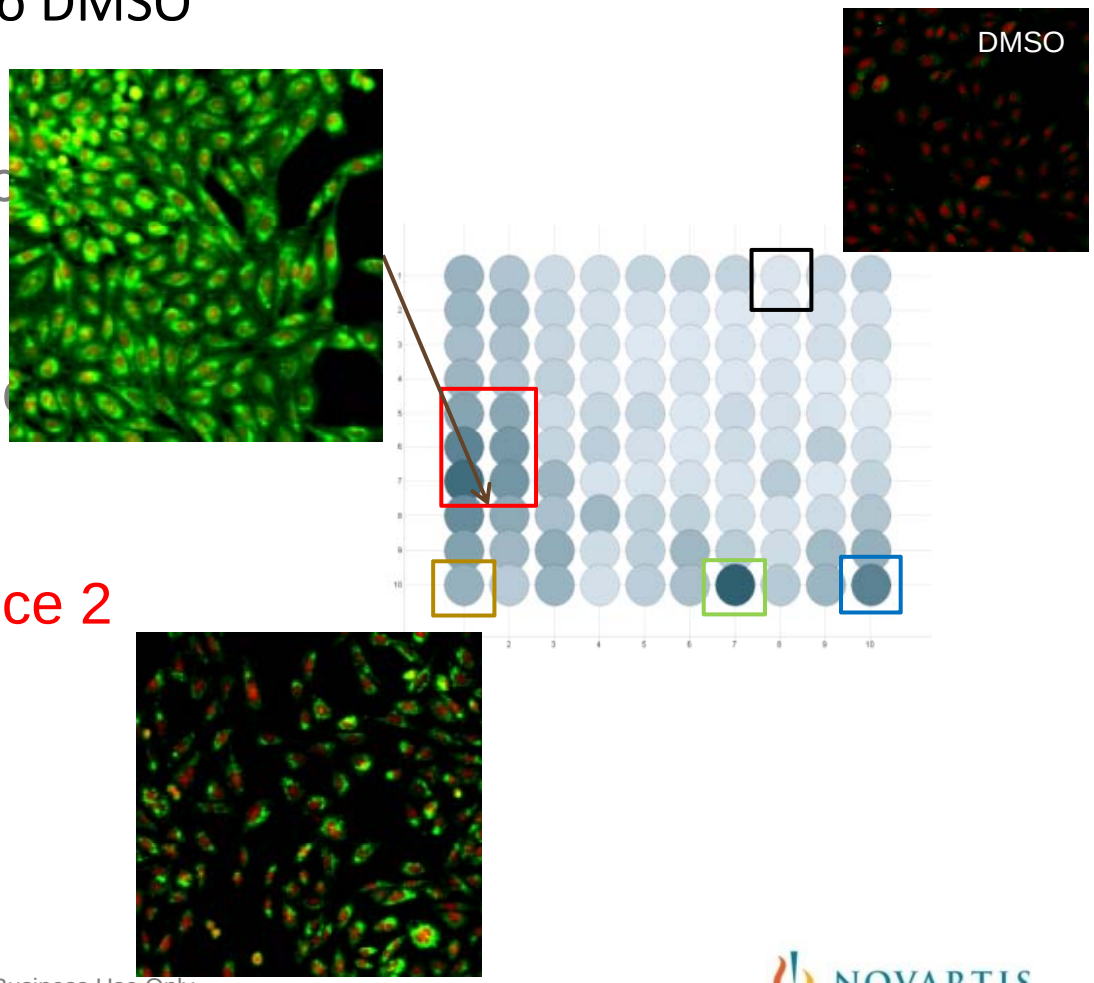
A compound similarity search suggested targets involved in the cell cycle

NOVARTIS

„Hit“ exploration based on SOM

High Mahalanobis distance 2

- Exploration of groups that have remarkable Mahalanobis distance: the darker the more distant to DMSO



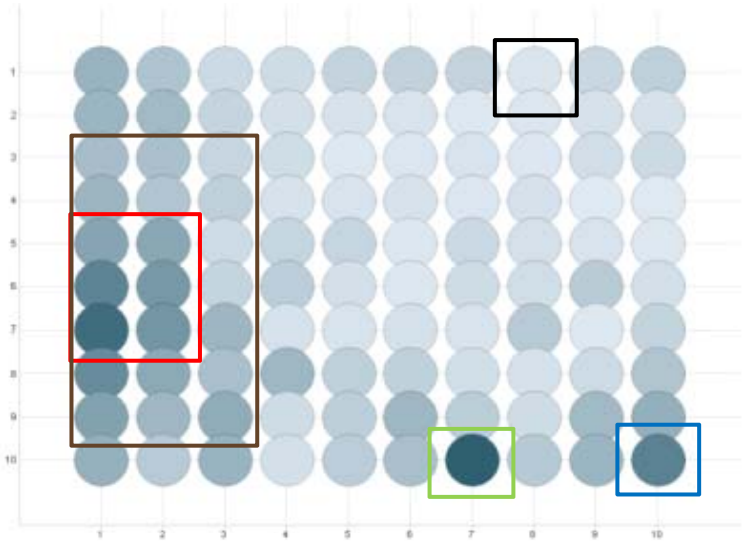
- Low Mahalanobis distanc
- Single high Mahalonobis
- High Mahalanobis distan
- Group of 8 compounds
- High Mahalanobis distance 2

Very similar to positive control = potential hits

„Hit“ exploration based on SOM

Suggestion of additional hit candidates

- All validated hits in neighboring groups around the clusters with high Mahalanobis distance
- Clusters with similar profiles are close to each other
- Neighboring groups also with high Mahalanobis distances



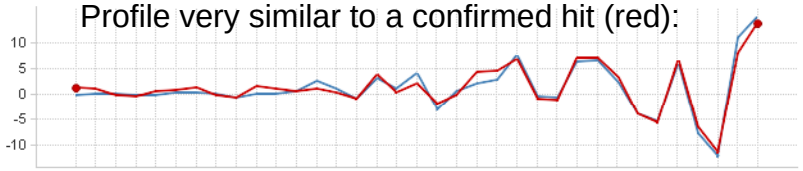
✓ All 31 confirmed compounds selected by univariate analysis were found

✓ 19 additional compounds were identified

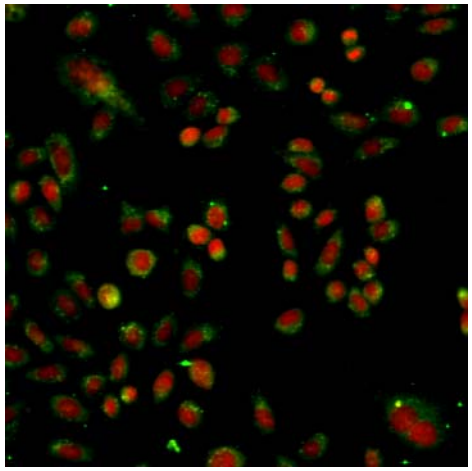
„Hit“ exploration based on SOM

19 additional hits selected

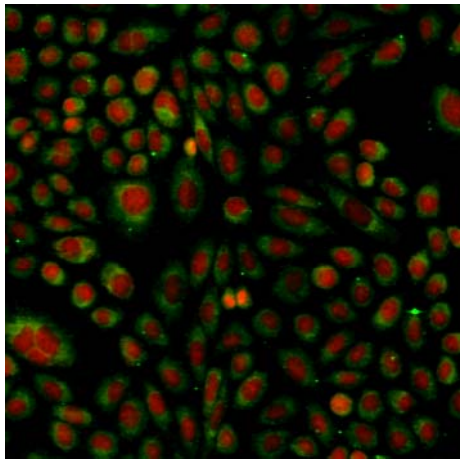
✓ This compound has a very similar profile to a compound which was identified and confirmed with univariate analysis, however a different chemical scaffold



Selected compound 6



Confirmed compound 4



Novel chemical scaffold

NOVARTIS

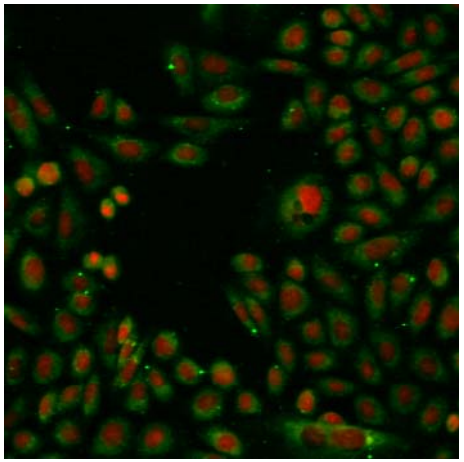
„Hit“ exploration based on SOM

19 additional hits selected

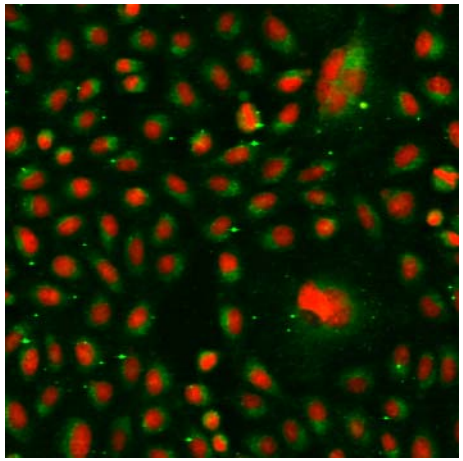
✓Similar chemical scaffold to validated hit identified

✓These compounds were not selected being slightly below the threshold setting with univariate analysis

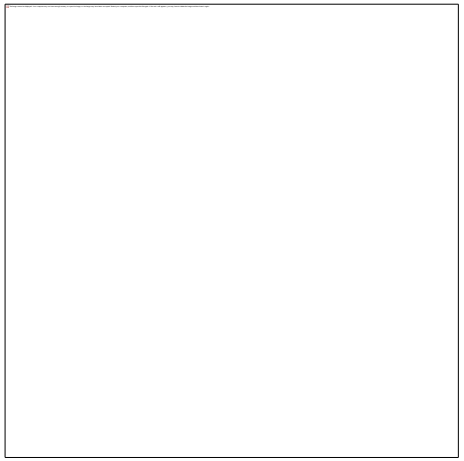
Selected compound 7



Selected compound 8



Confirmed compound 5

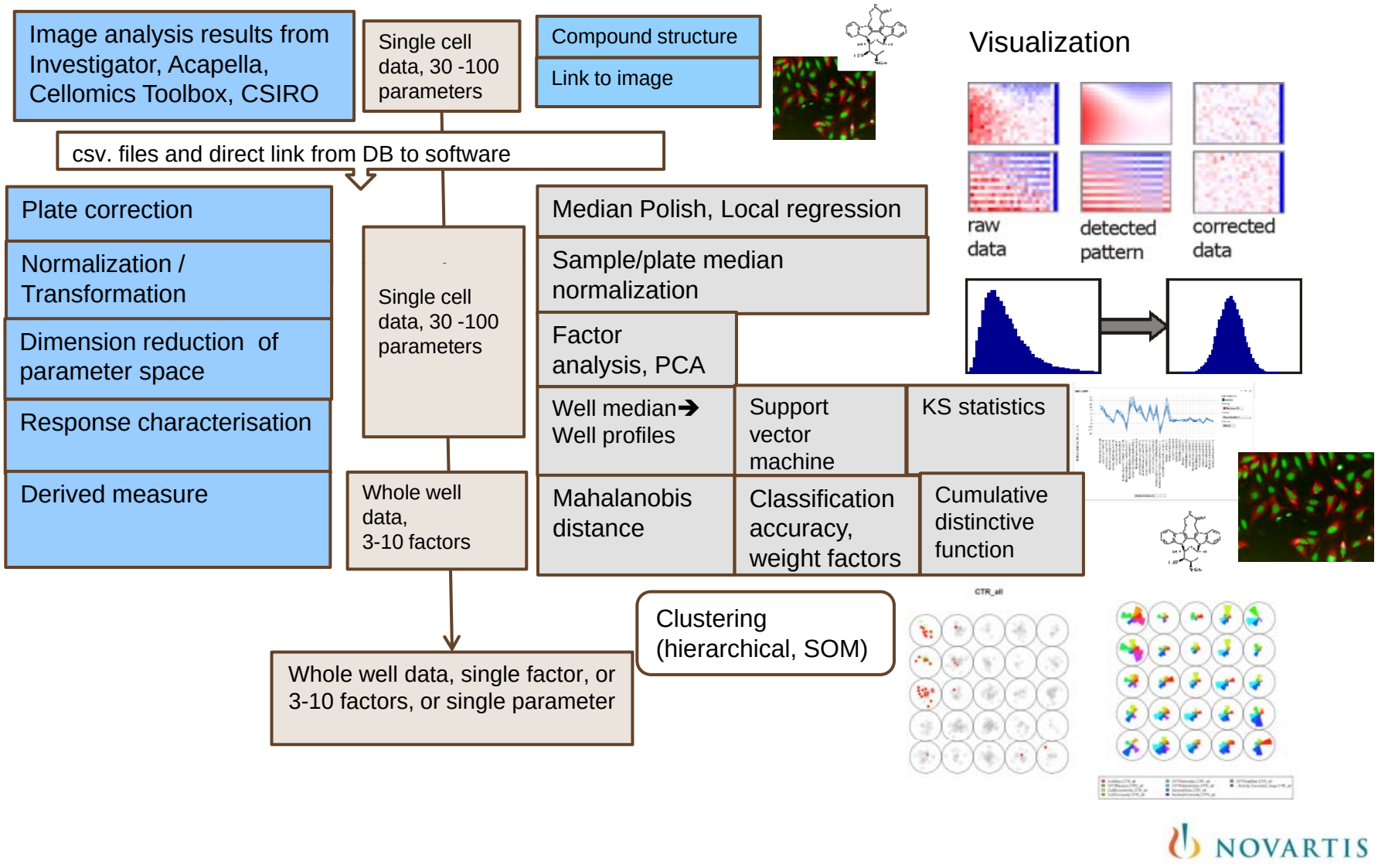


Chemical scaffold similar to confirmed hit

NOVARTIS

Multi-parametric data handling

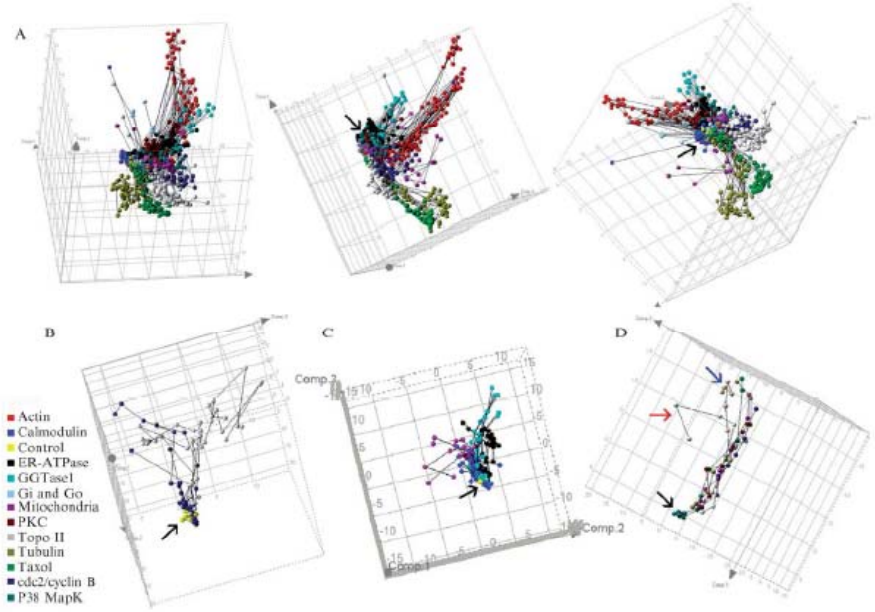
Data processing pipeline

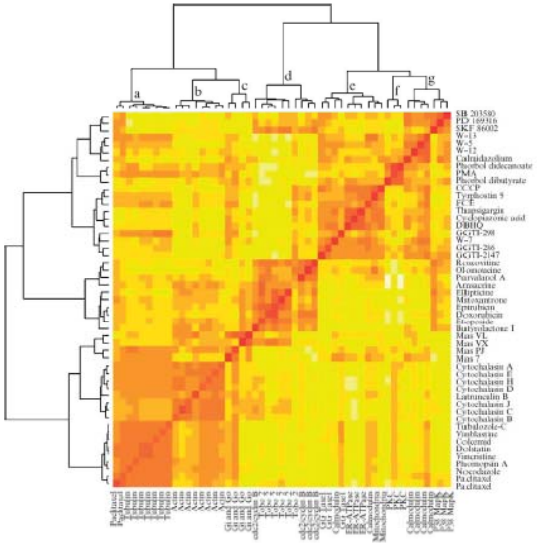


NOVARTIS

Dose response analysis

Adams CL et al. Meth Enzym 2006, Vol 414, p 440





Principal component analysis:

Lines connect increasing concentrations of a single compound in a single well. Concentration–response curves are colored by mode of action

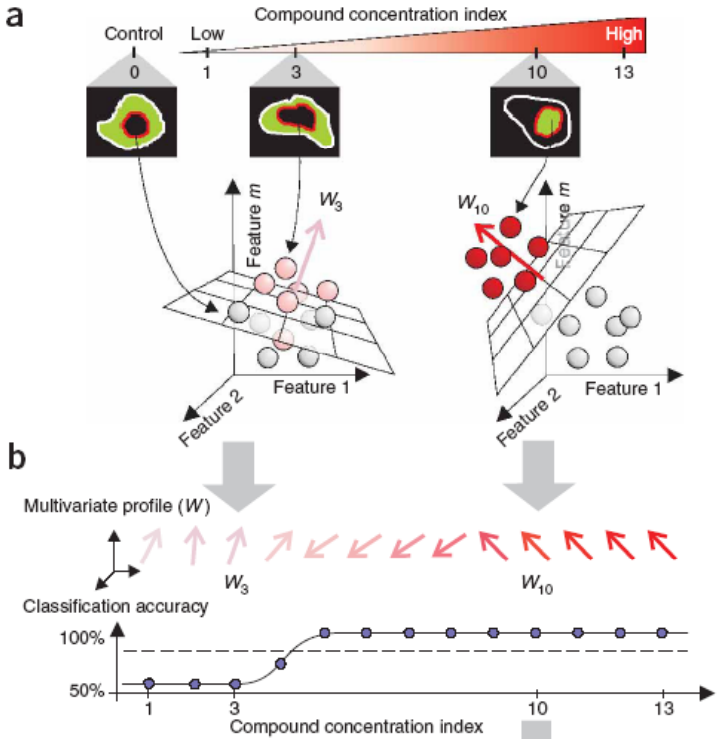
Clustering using angle dissimilarity measure:

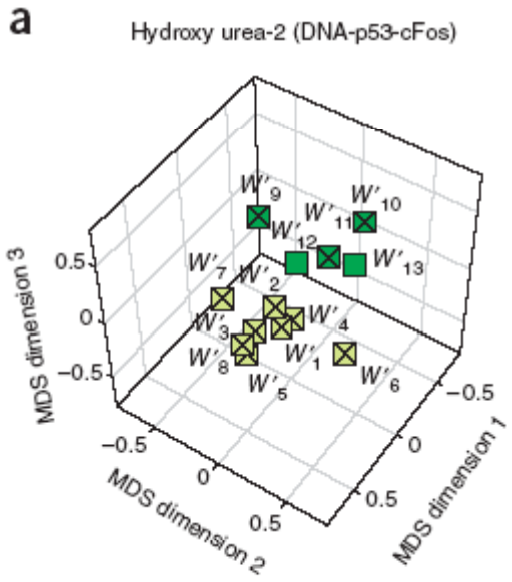
- This measure aligns multi-variate dose-responses by their potencies or distances from the control.
- Heat plot showing correlation between the angle dissimilarity measure

NOVARTIS

Dose response analysis

Loo L-H et al. Nature Meth 2007, Vol 4, p 445





Support vector machine analysis:

•Determined the optimum hyperplane separating cells into treated and untreated classes for each compound concentration (white hyperplanes).

•Hyperplane orientations are specified by weight factors.

Titration clustering:

•The numbers of clusters were determined automatically using a clustering validation algorithm in the original feature space.

•Titration clustering grouped the weight factors of different concentrations into different clusters.

NOVARTIS

Summary and Conclusions

- ✓ Groups of compounds with pronounced phenotypic changes were clustered together ➔ even bacterial infection was detected
- ✓ Additional compounds compared to univariate analysis were found
  - ✓ compounds with chemical scaffolds similar to validated hits
  - ✓ compounds with novel chemical scaffolds were detected
- ✓ Hit selection criteria have to be selected individually
- ✓ Selection of compounds is not restricted to a specific readout therefore less biased - however, for a specific readout it could be less sensitive

Acknowledgements

- Anne Kümmel

- Christian Parker

- Paul Selzer

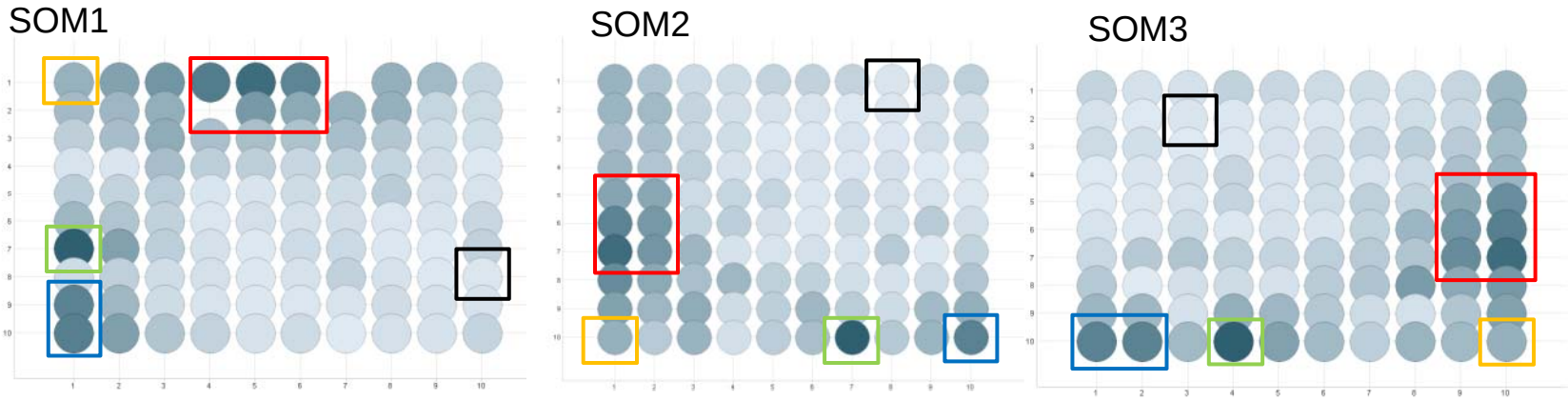
- Daniela Siebert

Hanspeter Gubler
Martin Beibel
Nicolas Fay
Marjo Götte
Yvonne Ibig-Rehm
Jürgen Reinhardt
Peter Fürst

Backup

„Hit“ exploration based on SOM

- Exploration of groups that have remarkable Mahalanobis distance (the darker the more distant to DMSO)



- Low Mahalanobis distance
- Single high Mahalonobis distance
- High Mahalanobis distance region 1
- Group of 8
- High Mahalanobis distance region 2

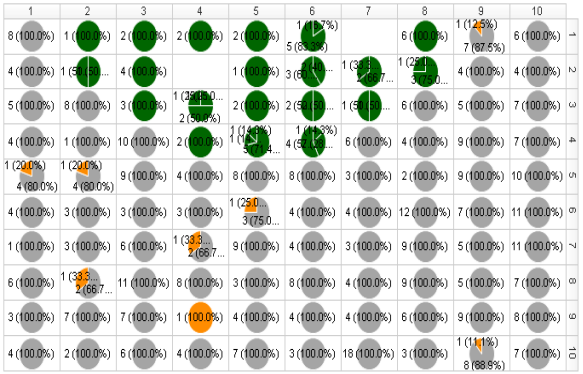
NOVARTIS

„Hit“ exploration based on SOM

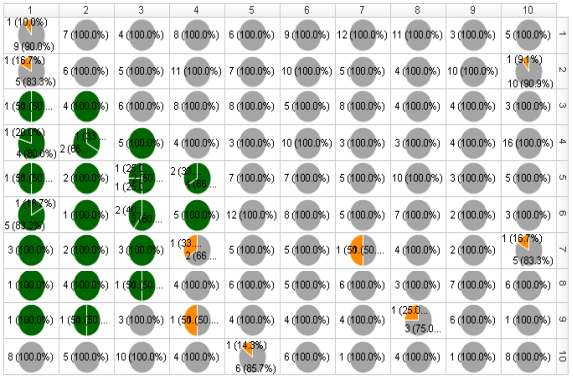
Suggestion of additional hit candidates

- All validated hits in neighboring groups
- Neighboring groups with high mahalanobis distances

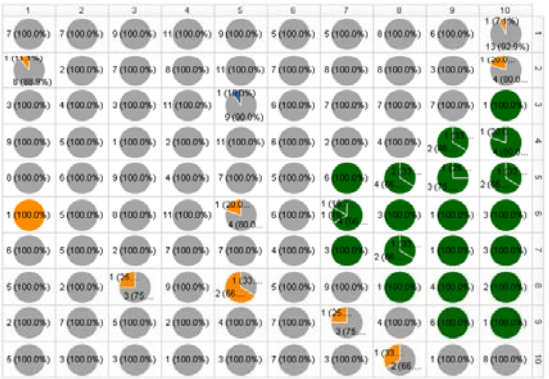
SOM1



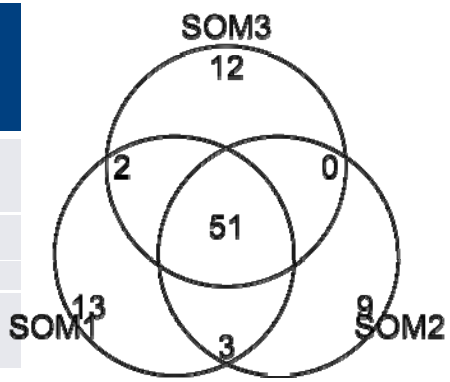
SOM2

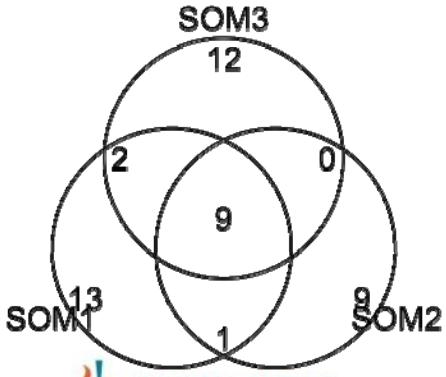


SOM3



cluster run	picked hits	picked InCell hits	picked valid hits	additionally picked hits
SOM1	69	44	31	25
SOM2	63	44	31	19
SOM3	65	42	30	23





NOVARTIS