Society of Biomolecular Screening
Geneva, Switzerland
Special Interest Group
September 15, 2005

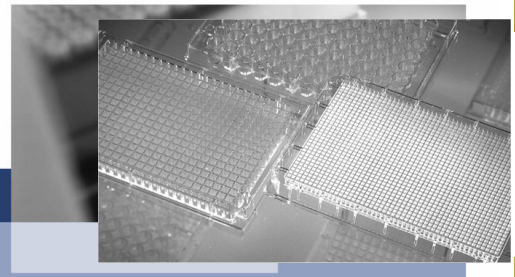
# Use of the AcuityXpress Informatics Platform For Expanding the Potential of High-Content Screening

Ralph Garippa, PhD
Research Leader
Roche Discovery Technologies

Damian Verdnik Ph.D.
Systems Analyst
Molecular Devices

Roche Discovery Technologies

# AcuityXpress Validation with Large Data Sets



- Analysis of Transfluor® - Evotec Technologies' Opera
  - 28,000 compounds
- Analysis of Transfluor - GE Healthcare's InCell 3000
  - 112,000 compounds
- Analysis of Transfluor - Cellomics' ArrayScan 3.1
  - 32 compounds, 10 point dose-responses curves
  - 5 cell lines
  - 3840 wells (10 plates)

- *All data generated at Hoffmann-La Roche, Inc., Nutley, New Jersey*

# aims of this study

## Evaluating multiparametric analysis for HCS

- Evaluate data from multiple HCS platforms
  - Evotec Opera
  - GE Healthcare INCell 3000
  - Cellomics ArrayScan 3.1
- Evaluate data pre-processing methods
  - Plate-specific normalization
  - Corrections for other plate-specific effects
  - Other data transformations
- Evaluate multiparametric analysis algorithms
  - Principal components analysis
  - Self-organizing maps
  - Other (hierarchical clustering, gene shaving, gap statistic)
- Compare results with data analyzed manually

# AcuityXpress – analyze previously processed images

- **OPERA**
  - 28,000 compound screen
- **InCell 3000**
  - 112,000 compound screen
- **ArrayScan 3.1**
  - 32 compounds, 10 point dose-responses curves
  - 5 cell lines
  - 3840 wells (10 plates)

Image Processing
Via instrument specific algorithms

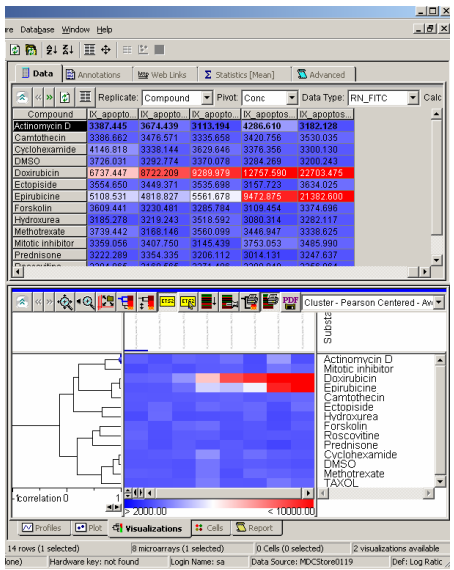
Hits

**IMPORT**
Multi-parameter data file

MDCStore Database

**AcuityXpress™**

Cellular Informatics
Quality Control
Data filtering
Visualization
**Hits by Profiling**



Roche Discovery Technologies

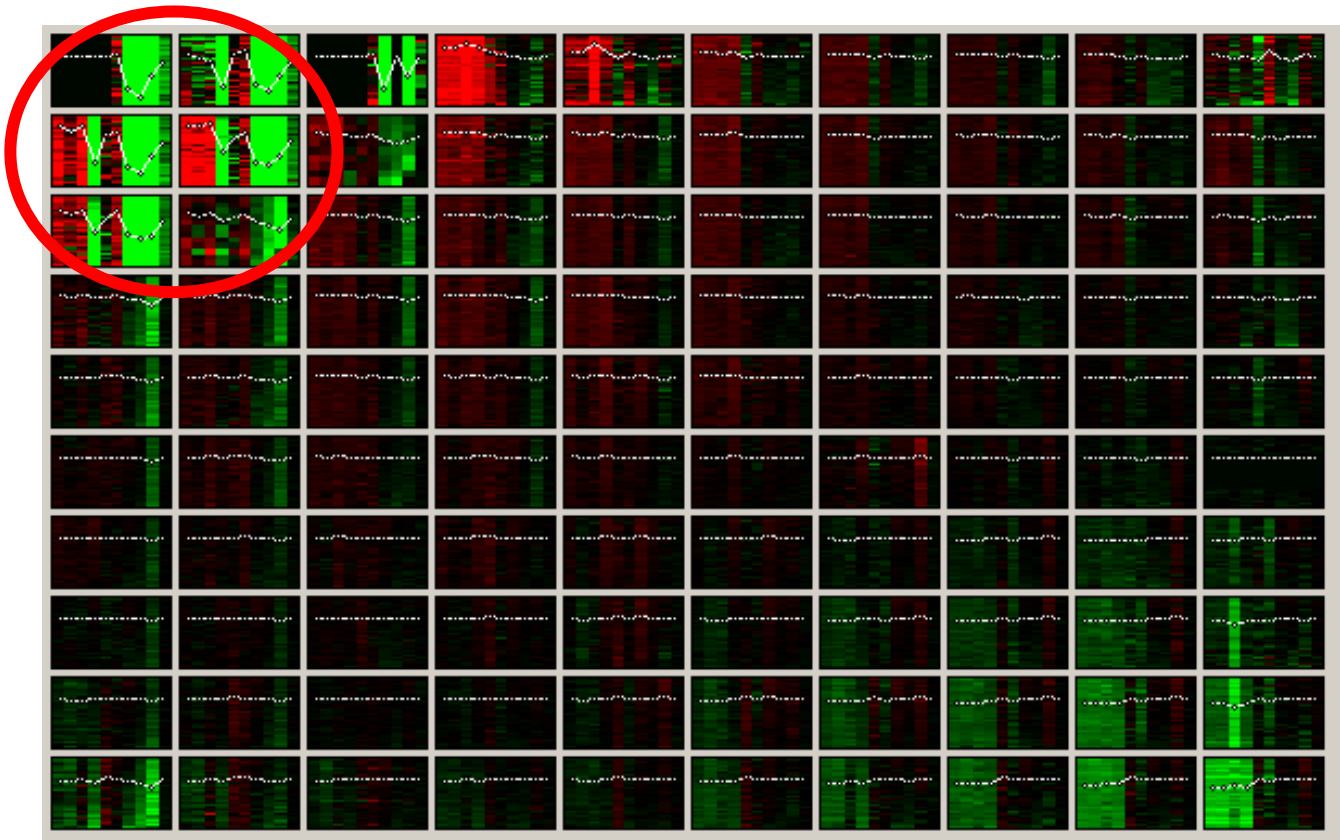
# Transfluor dataset 1

Acquired on the Evotec Opera

- 80 x 384-well plates
- 28,000 compounds
- Well-based statistics extracted using Evotec software and a customized "Roche GPCR" algorithm
- Data pre-processing
  - Normalize to negative controls on each plate
  - Log normalized results
- Use AcuityXpress to construct a dataset of 28,162 rows of compounds and 10 columns of cell measurements

# AcuityXpress Analysis: Opera

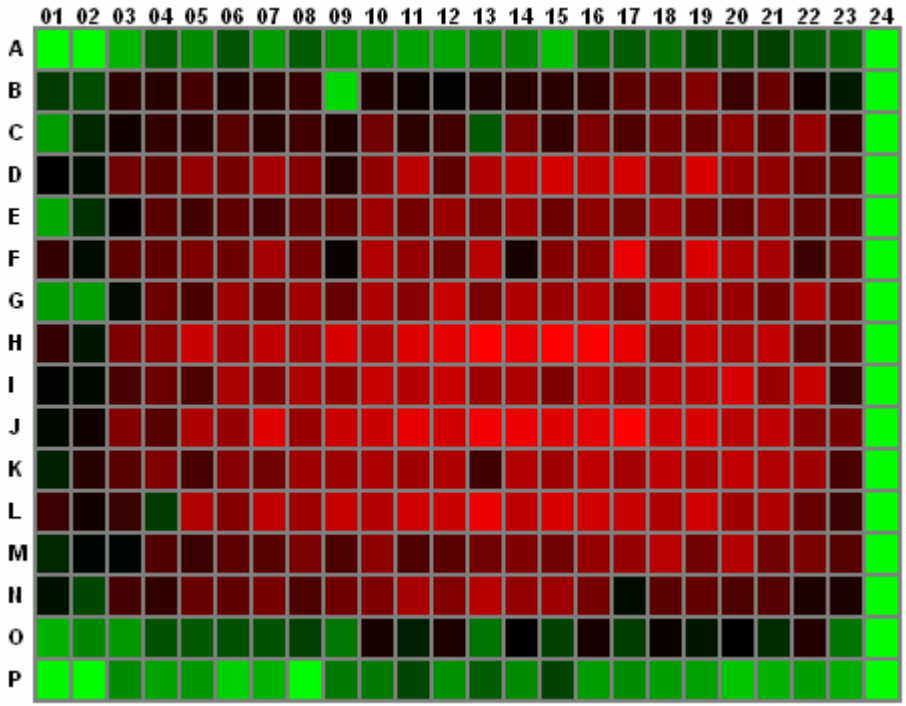
- Self-Organizing Maps
  - Analyzes total number of wells with anomalously low numbers of cells in a single image area: 30,720 well map (80 plates)



Roche Discovery Technologies

# AcuityXpress Analysis: Opera

- Investigation with the Plate view in AcuityXpress reveals edge effects. This is an average over all 80 plates showing variation in cell number per image:

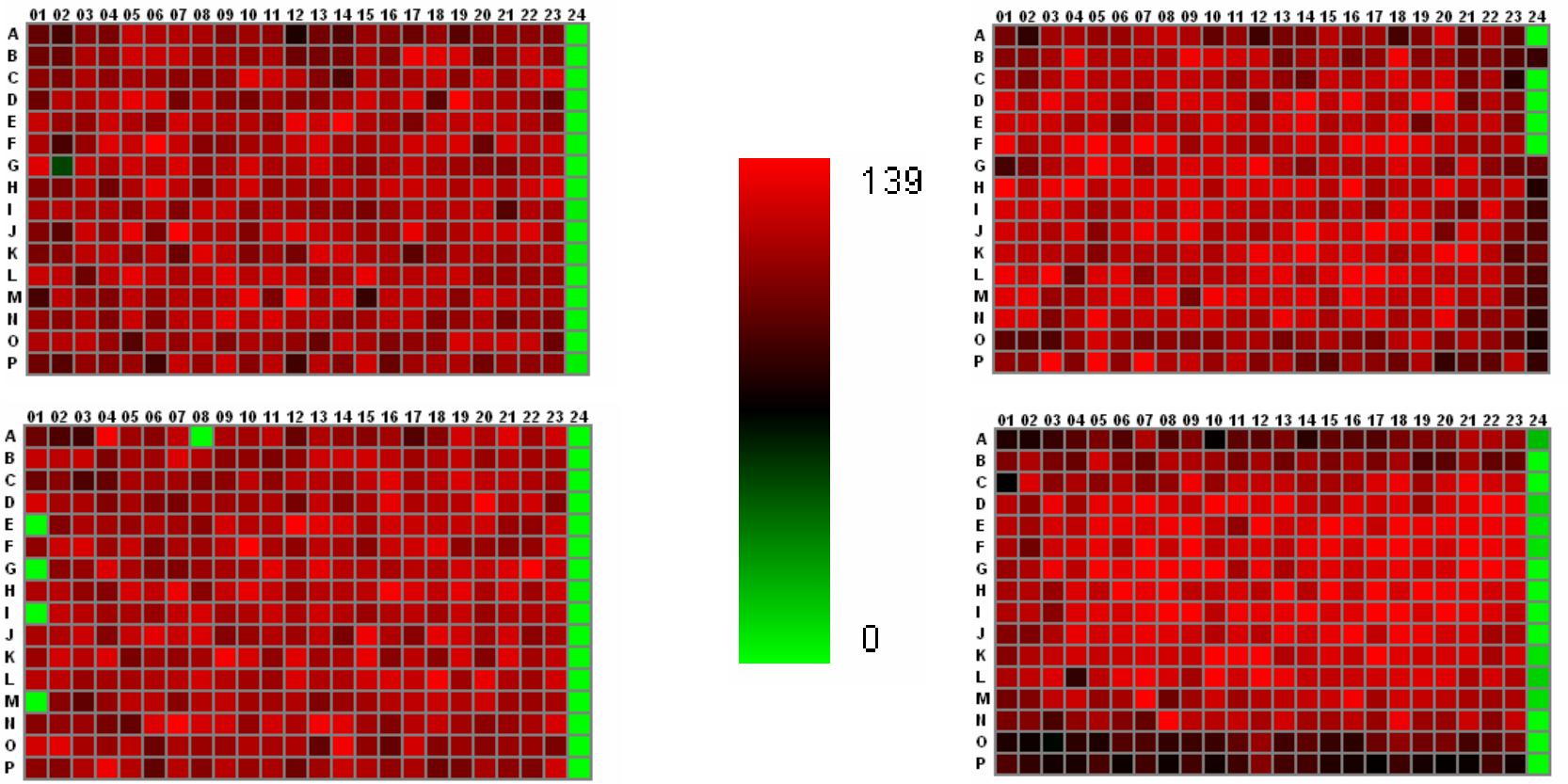



- Many wells in column 24 have < 5 cells per image
- Top and bottom rows also have low cell numbers

Roche Discovery Technologies

# AcuityXpress Analysis: Opera

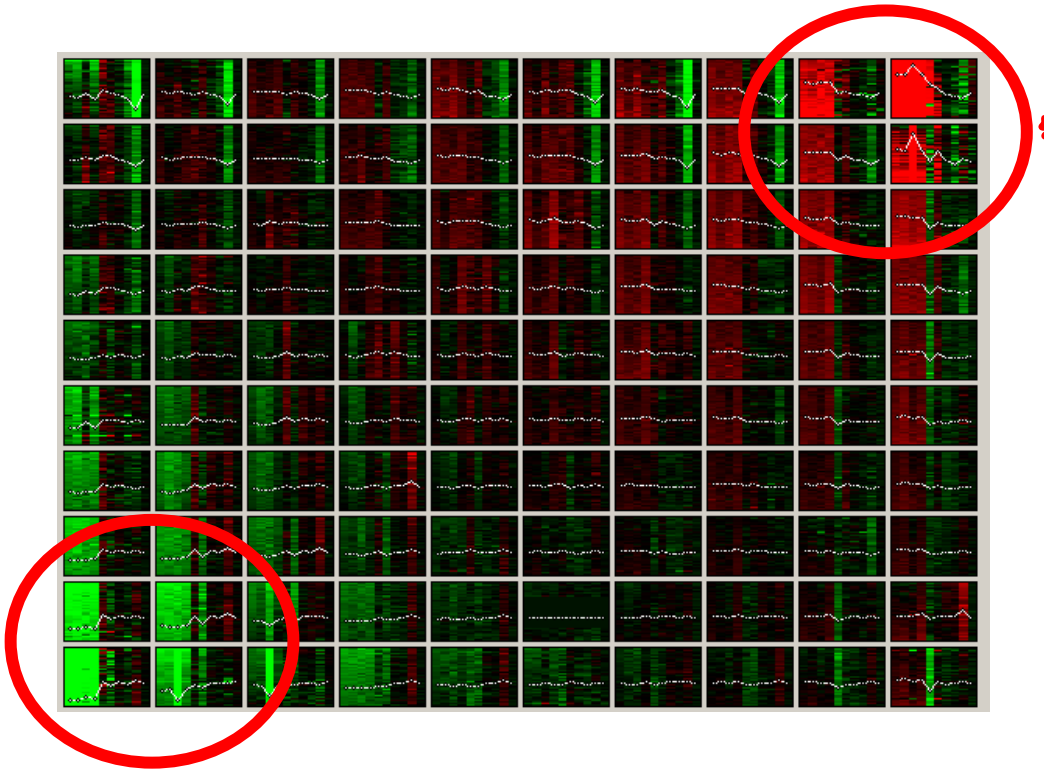
## A selection of plate events: cell number:




Roche Discovery Technologies

# AcuityXpress Analysis: Opera

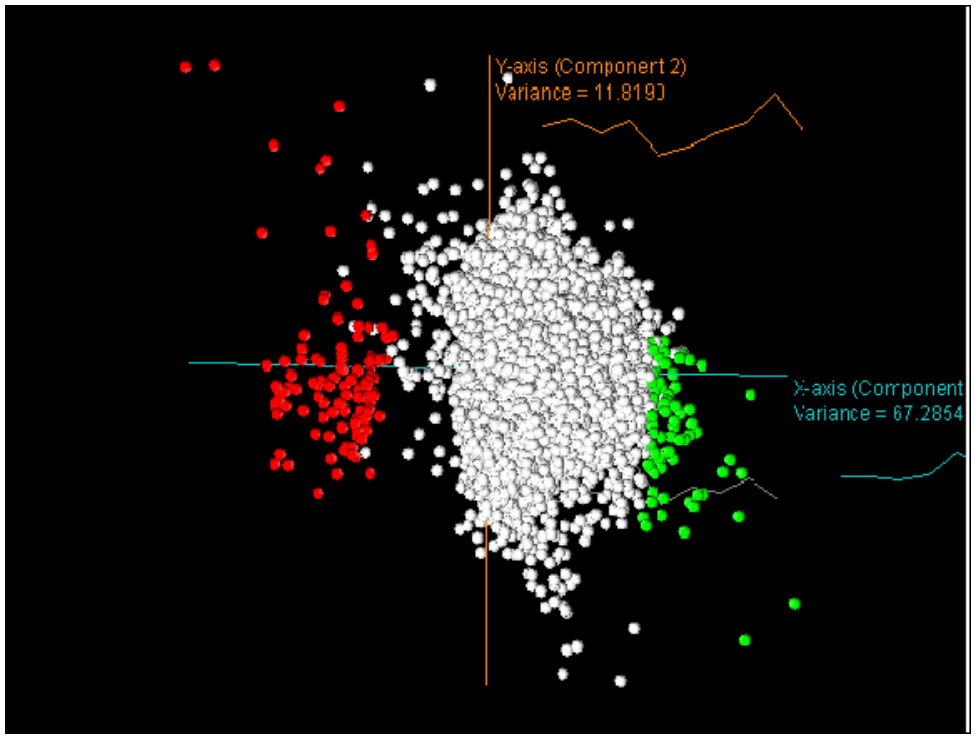
- Use a database query to exclude wells that have low cell numbers relative to the mean, and do another Self-Organizing Map:



- This reveals two compound-response profiles that are very different to the basal response

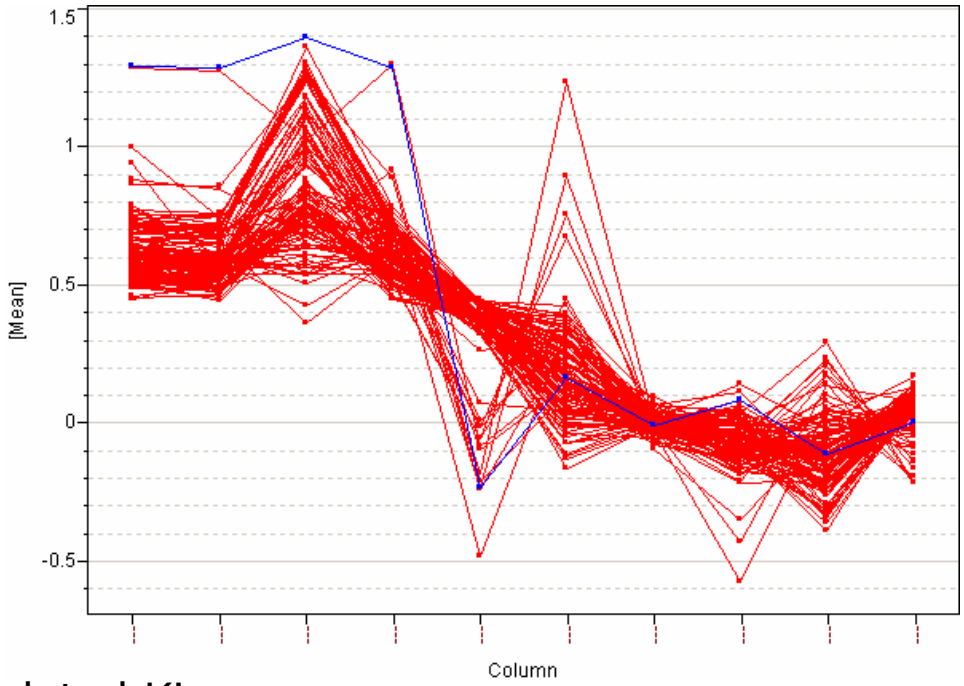
# AcuityXpress Analysis: Opera

- The two groups are consistent with the results of principal components analysis, where the **Red** and **Green** groups are shown as significant outliers on component 1:



# AcuityXpress Analysis: Opera

- Group 1 contains GRK and 114 other compounds with the same profile as GRK: high numbers of translocation centers relative to the control basal cells (GRK is blue)
- Filtered compounds with low cell number
- List of compounds to re-screen




GRK-G-protein Regulated Kinase

Roche Discovery Technologies

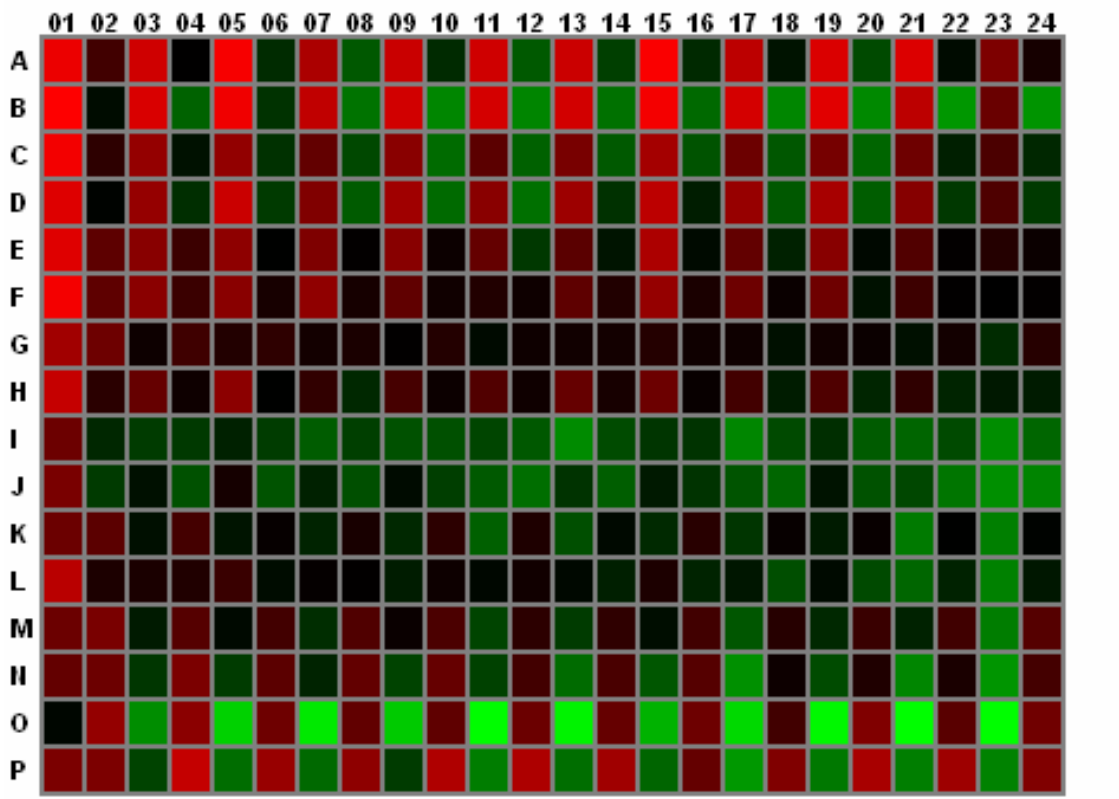
# Transfluor dataset 2

Acquired on the GE Healthcare's INCell 3000

- 352 x 384-well plates
- 112,642 compounds
- Well-based statistics extracted using instrument software and a customized "Roche GPCR" algorithm
- Data pre-processing:
  - Normalize to negative controls on each plate
  - Log normalized results
- Use AcuityXpress to construct a dataset of 112,642 rows of compounds and 14 columns of cell measurements

# AcuityXpress: InCell 3000

- AcuityXpress - Plate View
  - Indicates column effects in number of cells/well



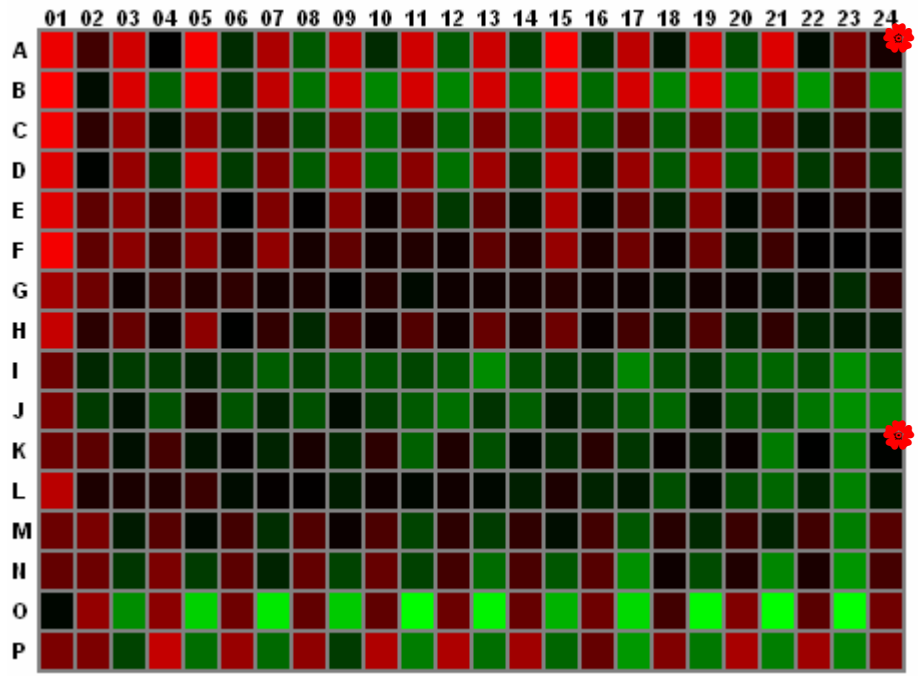

112,000 Compounds
352 plates

Roche Discovery Technologies

# AcuityXpress: InCell 3000

Acquired on the INCell 3000

- **Following the experience with the previous dataset, we look at cells per well in the Plate view, averaged over 352 plates:**

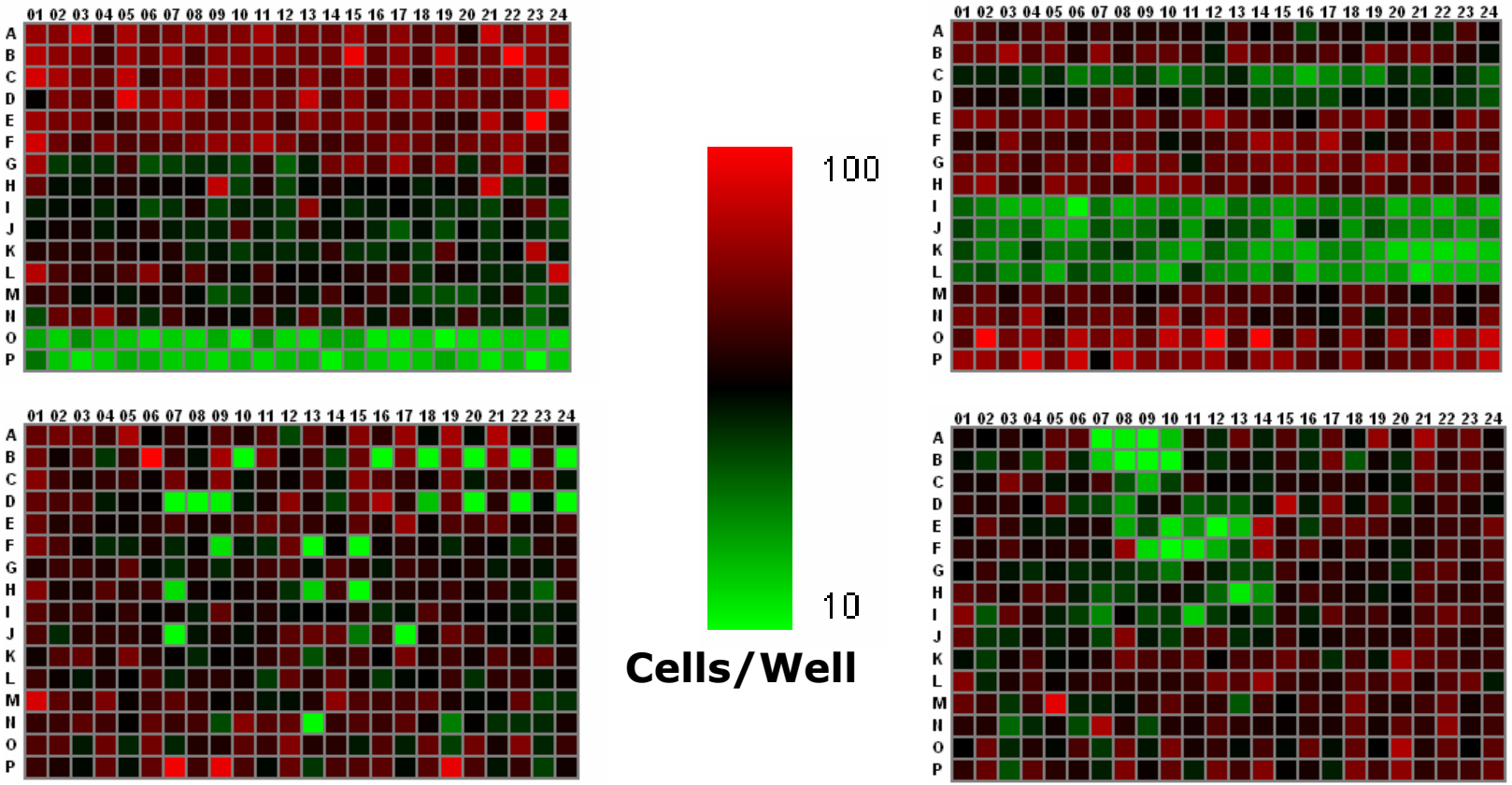



**Shows some interesting column effects for cell number, perhaps due to pipetting, but not at a significant level. However, individual plates show interesting effects.**

**Therefore, we filter the dataset so that only wells with more than 25 cells are included in the analysis.**

# AcuityXpress: InCell 3000

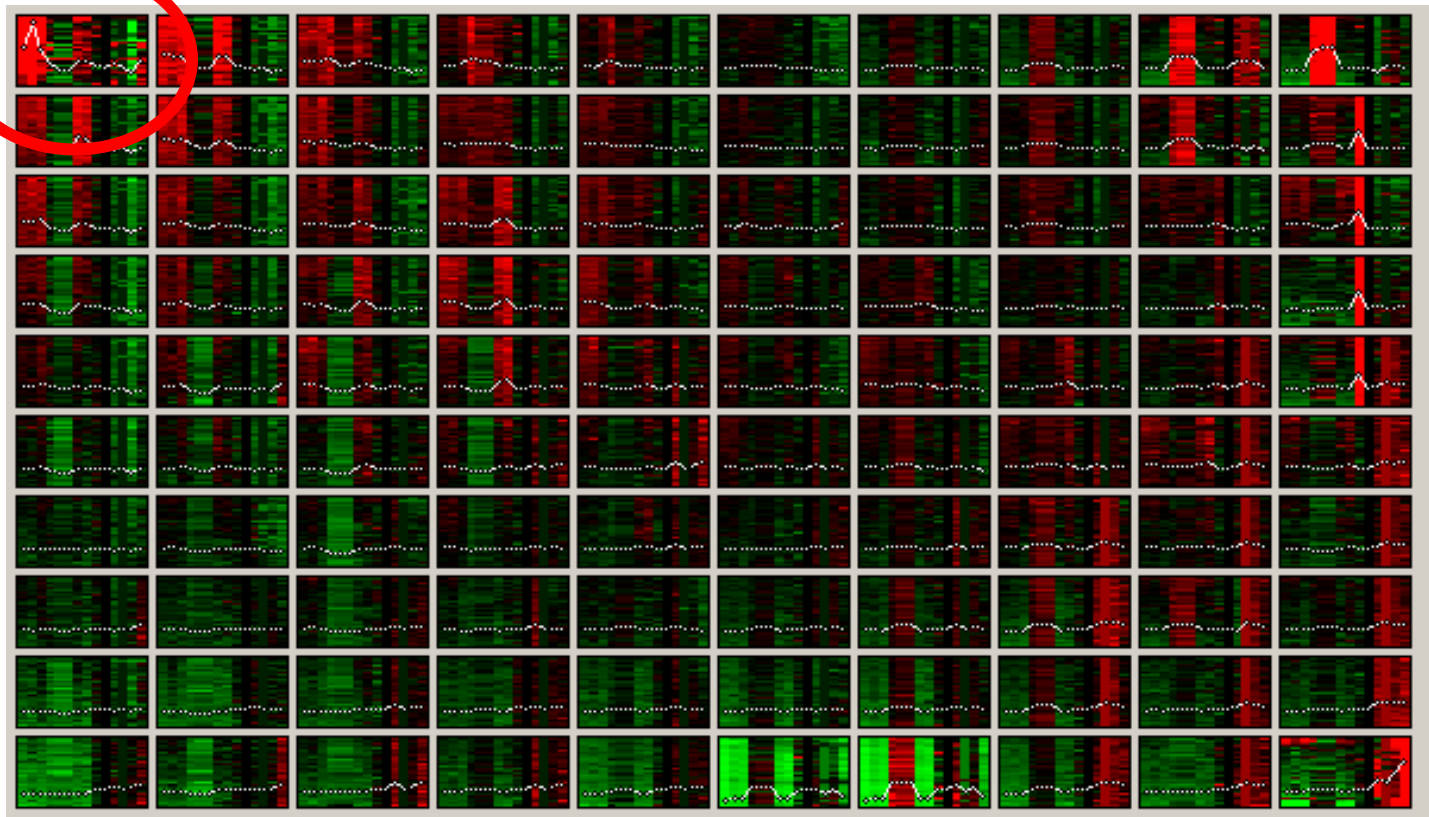
- Quality Control Features
  - Filtered dataset (only wells with > 25 cells are included)




Roche Discovery Technologies

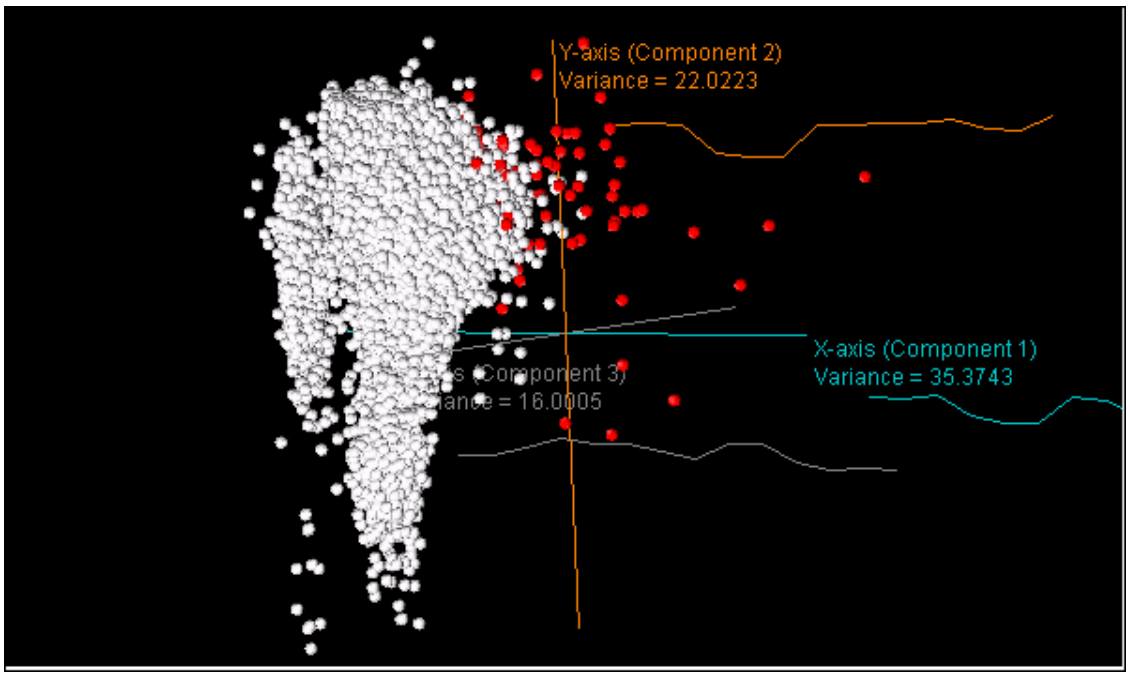
# AcuityXpress: InCell 3000

- Database Query
  - Exclude wells with low cell numbers relative to the mean, and to create another Self-Organizing Map



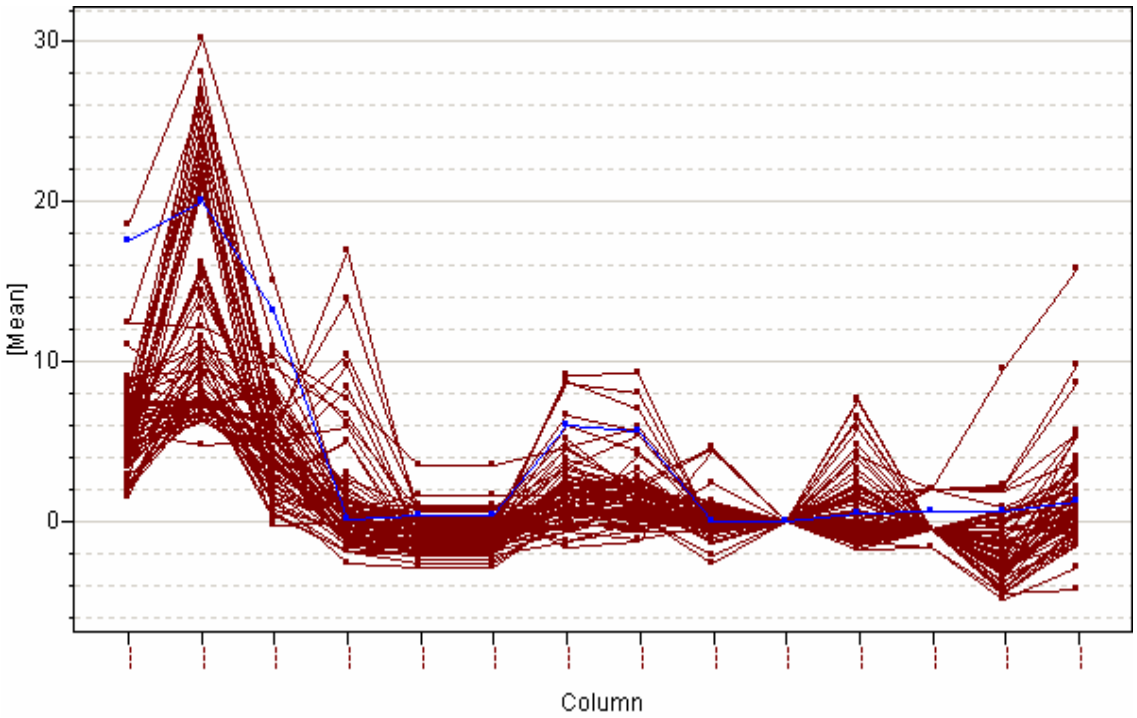
# AcuityXpress: InCell 3000

- AcuityXpress – Principal Component Analysis
  - This group is consistent with the results of PCA, where the group is shown as a significant outlier on component 1.
  - Red group is the significant outliers



# AcuityXpress: InCell 3000

- AcuityXpress – Profiling for Hits
  - Chart contains GRK (in blue) profile and 113 similar compounds
  - Ngrain (no of grains/cell),
  - Fgrain (average intensity of grains/cell) and
  - Agrain (average area of grains per cell) values higher than basal cells



# Transfluor dataset 3

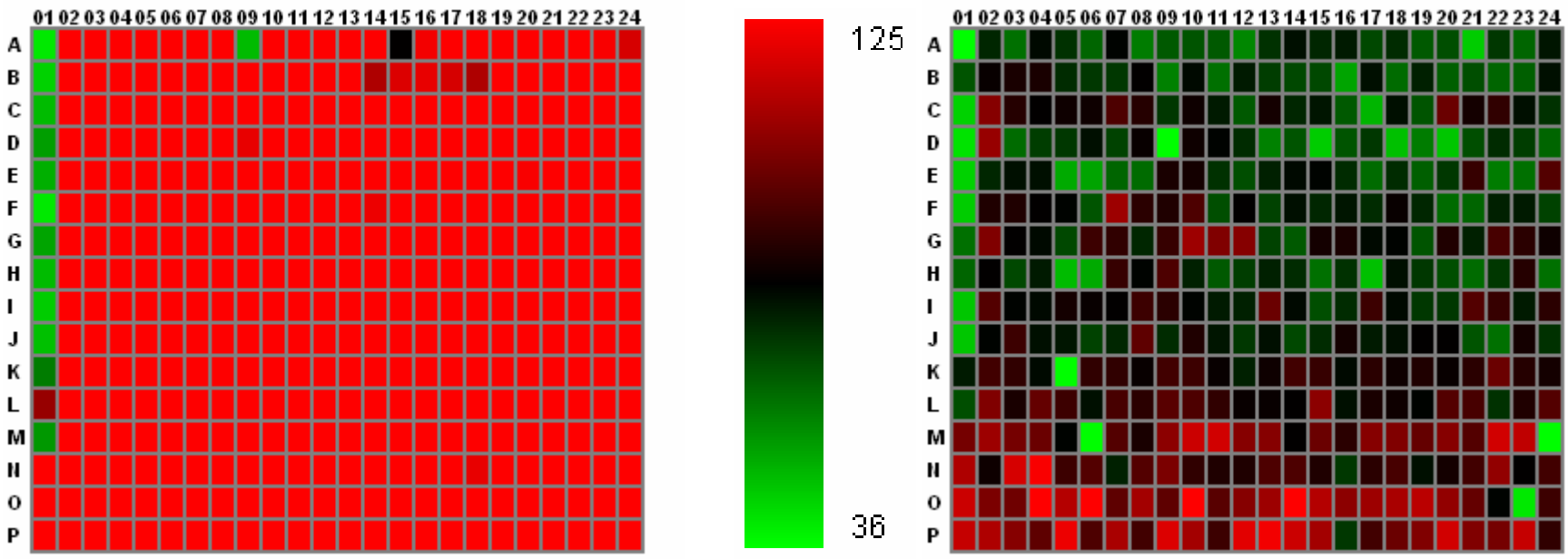
Acquired on the Cellomics ArrayScan 3.1

- 10 x 384-well plates, 2 plates for each of 5 cell lines (A-E)
- 32 compounds tested on the 5 cell lines
- 10-point dilution series from 1 nM to 30 μM in duplicate
- Extracted well-based statistics using Cellomics GPCR Bioapplication software
- No data pre-processing:
  - Analyze unnormalized dose-response data
- Use AcuityXpress to construct a dataset of 160 rows of compounds and 10 columns of cell measurements at different concentrations

# AcuityXpress: Cellomics ArrayScan

Acquired on the Cellomics ArrayScan 3.1

## A selection of plate events: cell number:

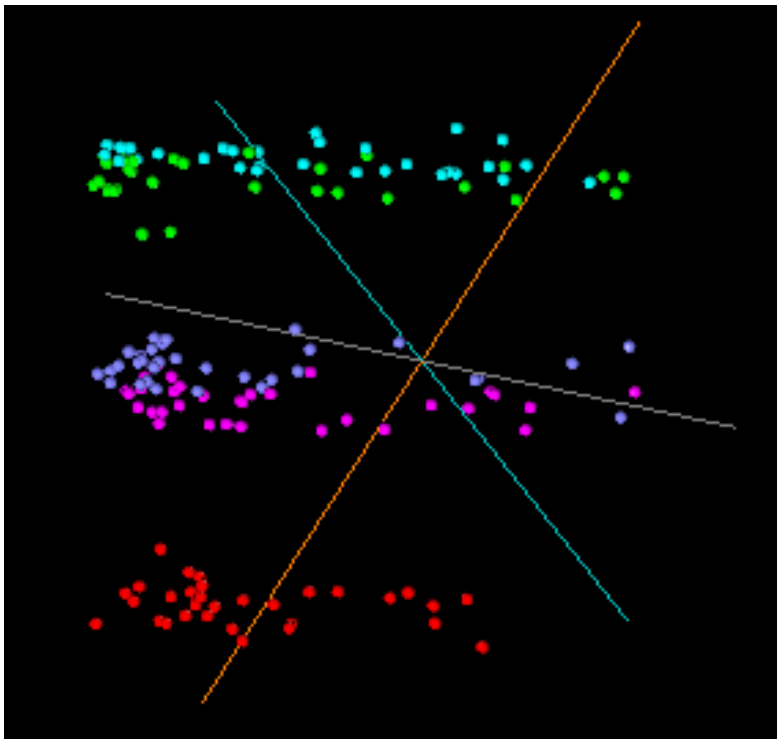


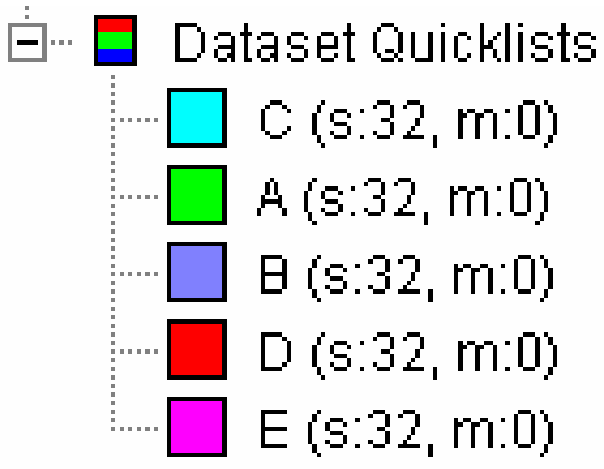

Roche Discovery Technologies

# AcuityXpress: Cellomics ArrayScan

Acquired on the Cellomics ArrayScan 3.1

- **Principal components analysis reveals the dose-response curves strongly segregated by cell line:**





# AcuityXpress: Cellomics ArrayScan

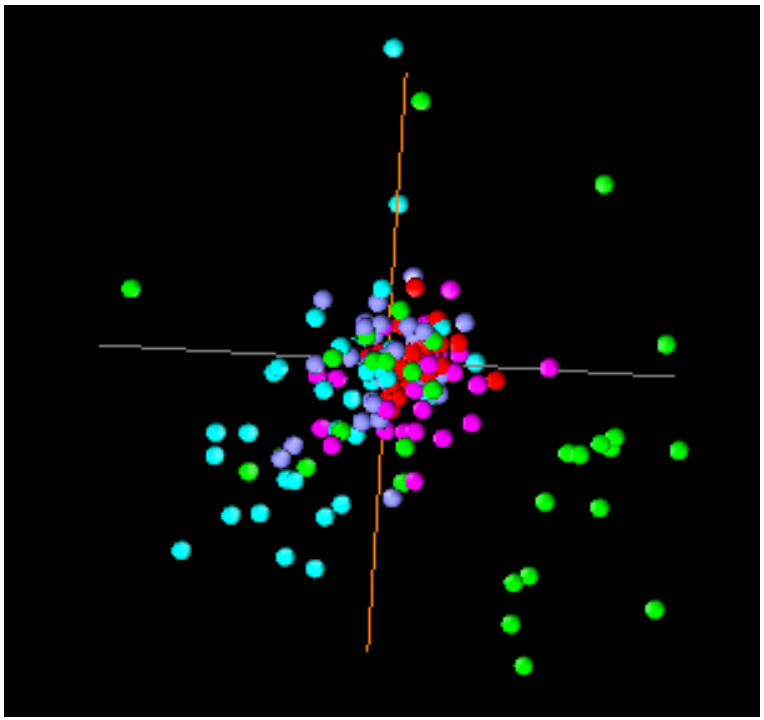
Acquired on the Cellomics ArrayScan 3.1

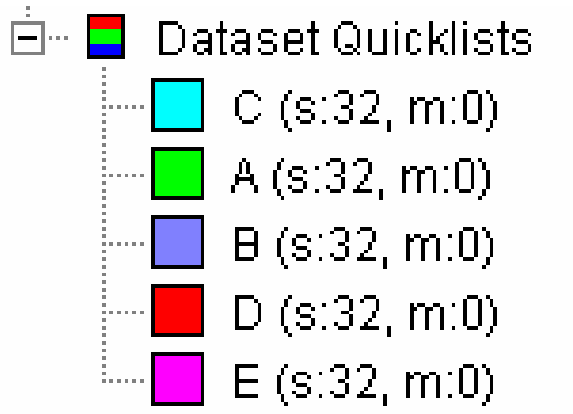
- We're interested in the effects of cell line on dose-response, so we need to remove plate-specific effects to be able to look at cell line effects only.
- We remove the plate-dependence by normalization to negative controls on each plate.

# AcuityXpress: Cellomics ArrayScan

Acquired on the Cellomics ArrayScan 3.1

- **Principal components analysis reveals a cell-line effect, but it is not as strong as in other studies. Cell lines A and C show the strongest dose-response effect (they are outliers):**

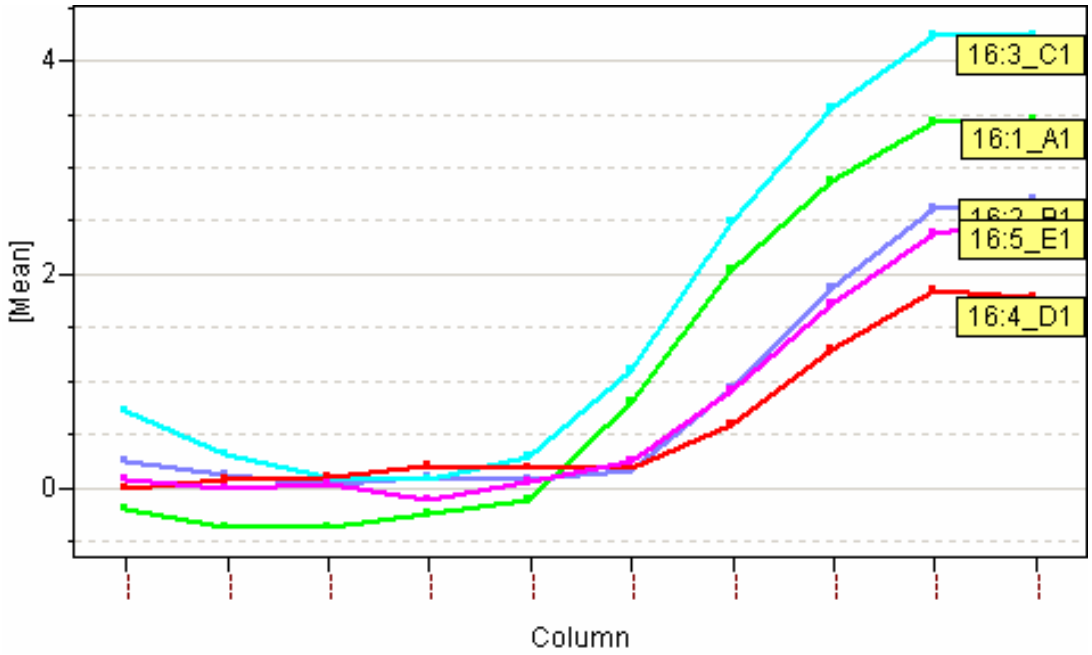




# AcuityXpress: Cellomics ArrayScan

Acquired on the Cellomics ArrayScan 3.1

- **Dose-response curves show the same variation in response by cell line (compound 16): simple line graphs**



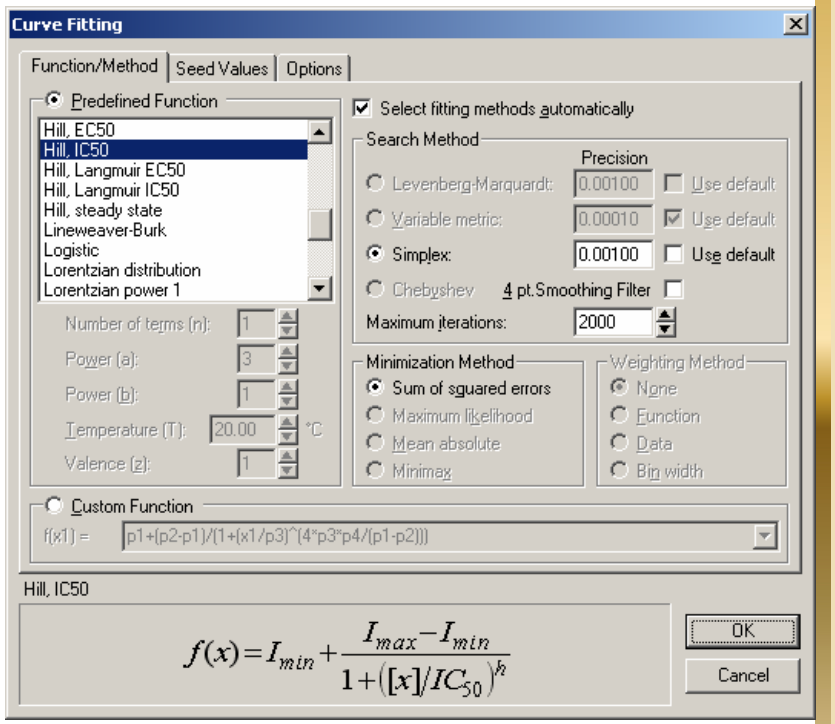

**Raw**

Roche Discovery Technologies

# AcuityXpress: Cellomics ArrayScan

calculate $IC_{50}$, $EC_{50}$

- AcuityXpress has sophisticated curve fitting functionality
  - Multiple fitting functions
  - Search methods
  - Minimization methods
  - Weighting methods
  - Seeding options
- Fitting of custom functions
- Model comparison
- Parameter fixing



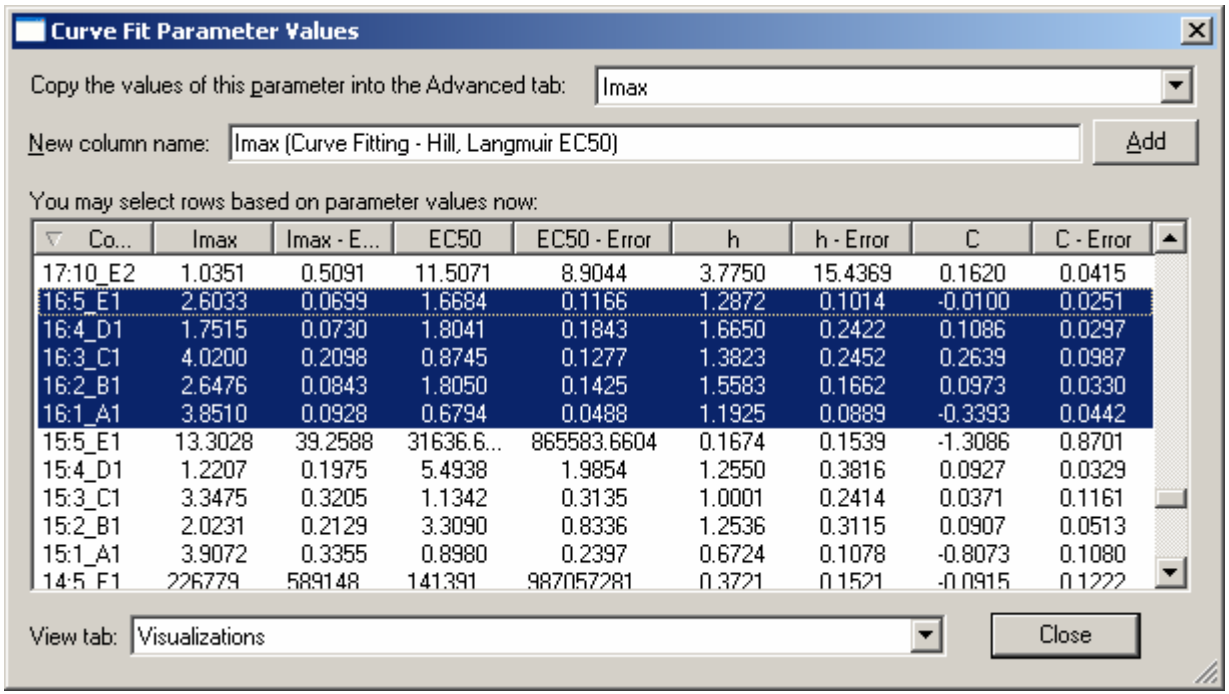

$$f(x) = I_{min} + \frac{I_{max} - I_{min}}{1 + ([x]/IC_{50})^{h}}$$

Roche Discovery Technologies

# AcuityXpress: Cellomics ArrayScan

## calculate $IC_{50}$, $EC_{50}$

- **AcuityXpress has sophisticated curve fitting functionality**
  - **Fit a whole experiment in one operation**
  - **Interactively view all fitted parameters for all fitted dose-response curves**

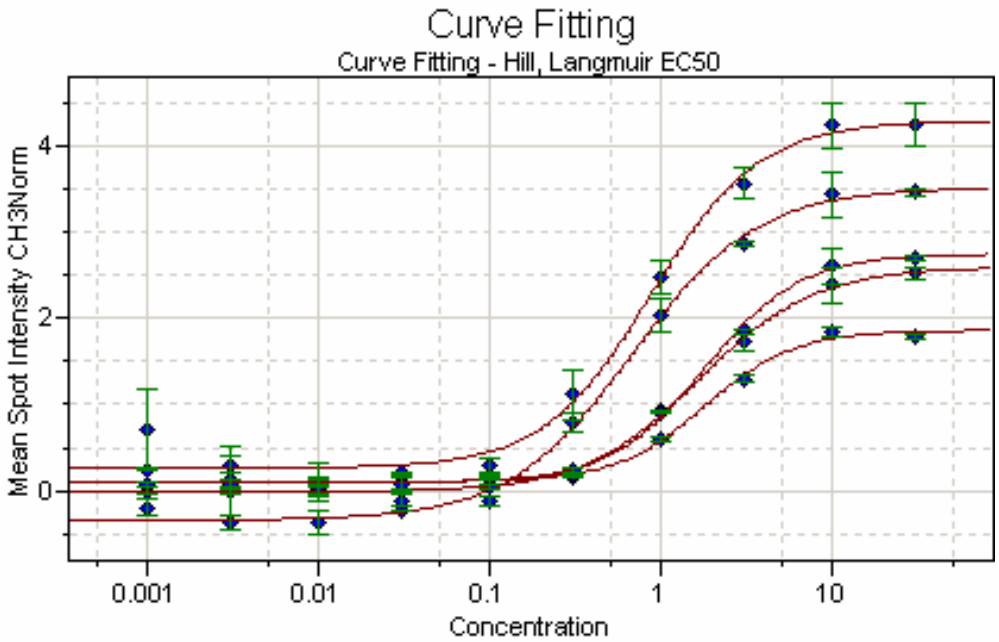



Roche Discovery Technologies

# AcuityXpress: Cellomics ArrayScan

Acquired on the Cellomics ArrayScan 3.1

- **Dose-response curves show the same variation in response by cell line (compound 16):**




**Fitted**

| Cell line | EC50 |
|---|---|
| C | 0.8745 |
| A | 0.6794 |
| B | 1.8050 |
| E | 1.6684 |
| D | 1.8041 |

Roche Discovery Technologies

# AcuityXpress: Cellomics ArrayScan

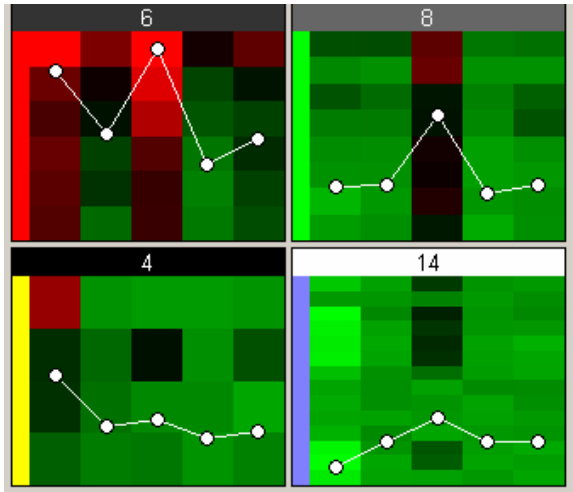
Acquired on the Cellomics ArrayScan 3.1

- Now configure the dataset to look at compound performance by cell line, independent of concentration
- 32 rows of compounds, 5 columns of cell lines (A-E)

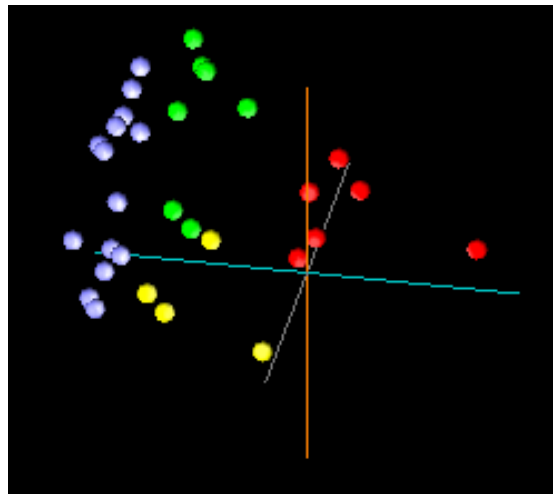
# AcuityXpress: Cellomics ArrayScan

Acquired on the Cellomics ArrayScan 3.1

- The compounds segregate into roughly 4 groups:
  - Strong responders on cell lines A and C (red)
  - Responders on C only (green)
  - Responders on A only (yellow)
  - Weak responders (mauve)




SOM



PCA

# AcuityXpress Conclusions:

- All instrument platforms produce data with easily identifiable effects
- Normalization to plate-specific negative controls is essential to remove plate-specific effects
- Other artifacts need to be monitored (e.g. cells/per well) and removed before advanced analysis
- Principal components analysis and self-organizing maps are useful and efficient methods for investigating HCS datasets:
  - Identify interesting subpopulations of compounds, whether by multiparametric analysis of single-point data or of dose-response
  - Identify variable performance of the Transfluor assay on different cell lines: magnitude of response and compound-specificity of response
  - Need to incorporate a rigorous measure of statistical significance, however, the results of principal component analysis suggest that the effects discovered are statistically significant
  - Need to be validated by comparing results with those from manual analysis

# AcuityXpress Summary:

- Multiparametric algorithms are powerful methods for analyzing HCS datasets quickly:
  - Small datasets: 32 compounds
  - Large datasets: 112,642 compounds
  - Very large datasets: limited only by processor power and computer memory
- These analyses are starting to gain wider acceptance in the HCS community
- An integrated toolbox like AcuityXpress may increase HCS productivity

# Acknowledgements


- Lee Babiss
- David Mark
- Rob Goodnow
- Ann Hoffman
- Anthony Aglione
- Shannon Hamilton
- Grazyna Kuryklo
- George Higgins
- Theresa Truitt
- Waleed Danho
- Mitch Martin
- Jim Rosinski
- Leon Garfinkel
- Keith Groco
- Dan Weiss
- Catherine Mccaffery
- Michael Good

**Molecular Devices:**

- Mike Sjaastad
- Paula Rickert
- Damian Verdnik

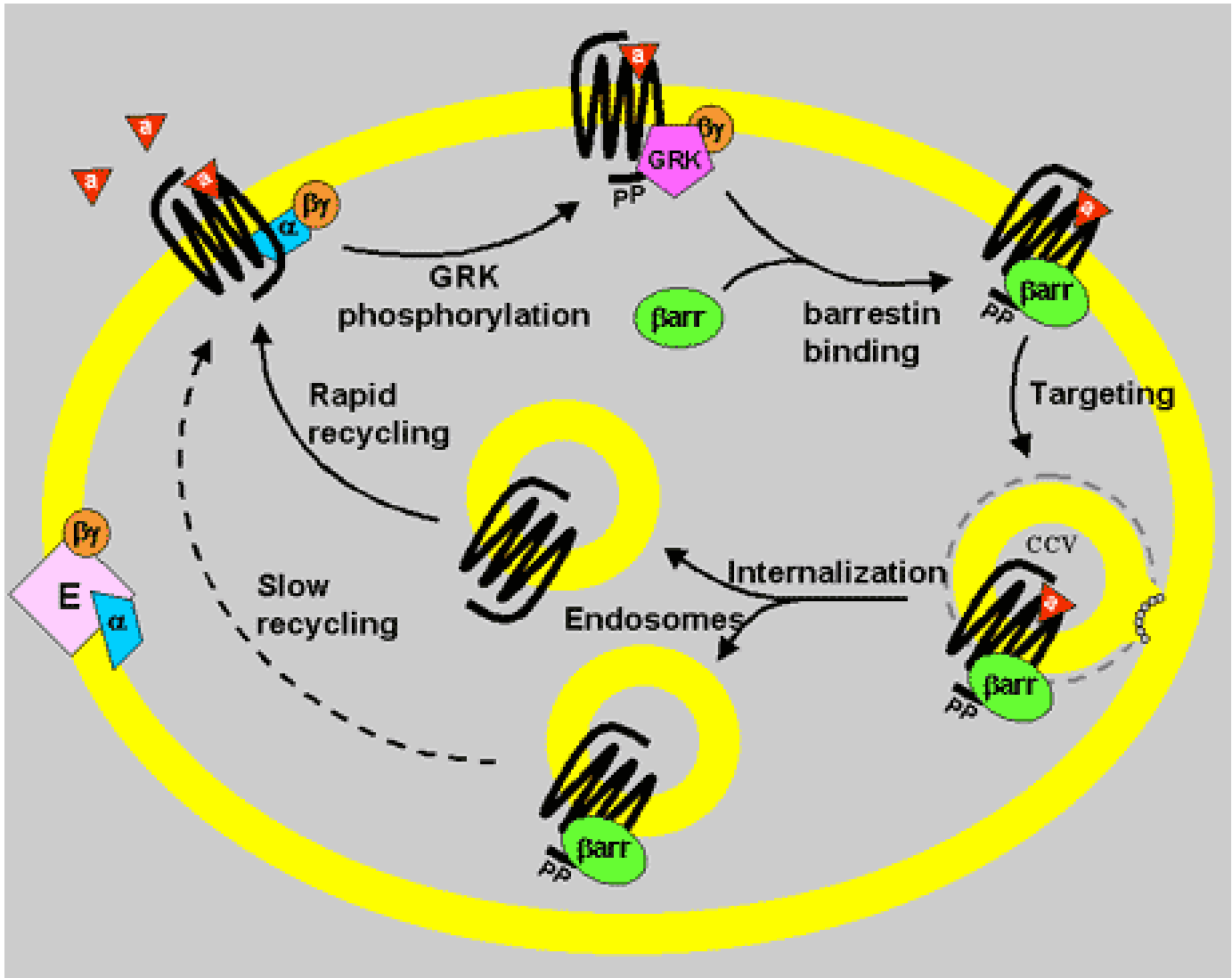
**Evotec Technologies**

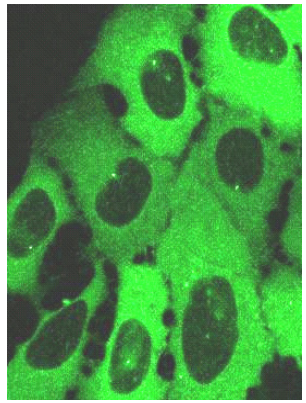
- Robert Rodewald
- Achim Kirsch
- Dennis Knapp

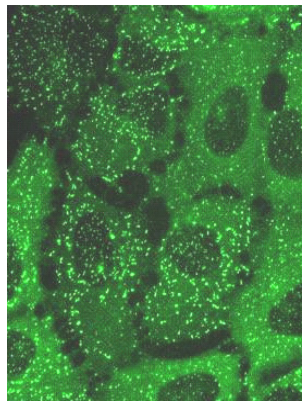
**Norak Biosciences -Xsira Pharmaceuticals**

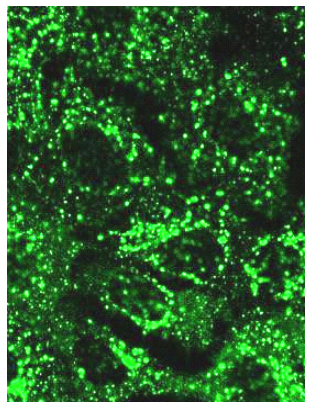
- Carson Loomis
- Robert Oakley
- Christine Hudson
- Allen Eckhardt
- Wen-Ji Chen

# The GPCR Signaling Cycle



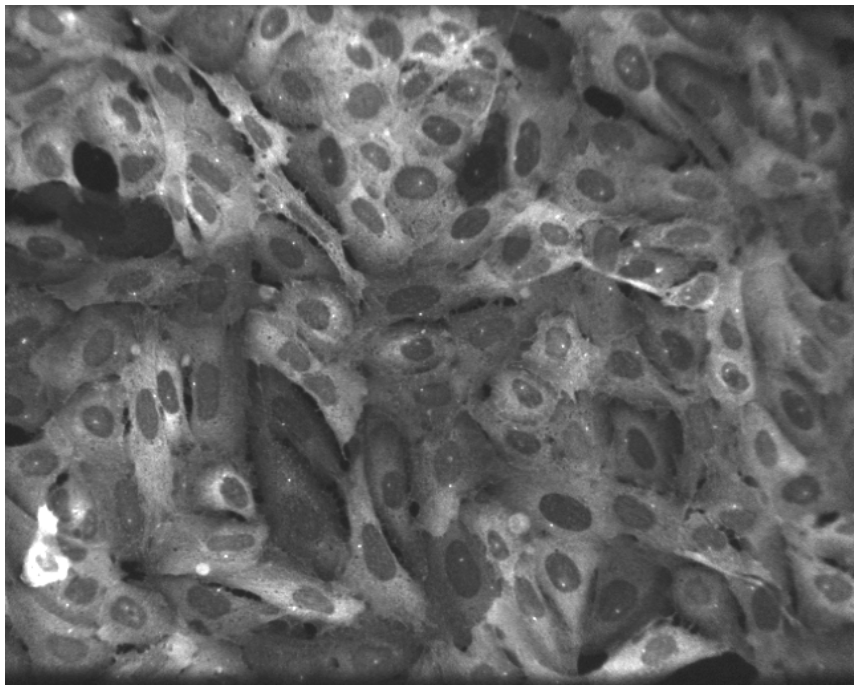





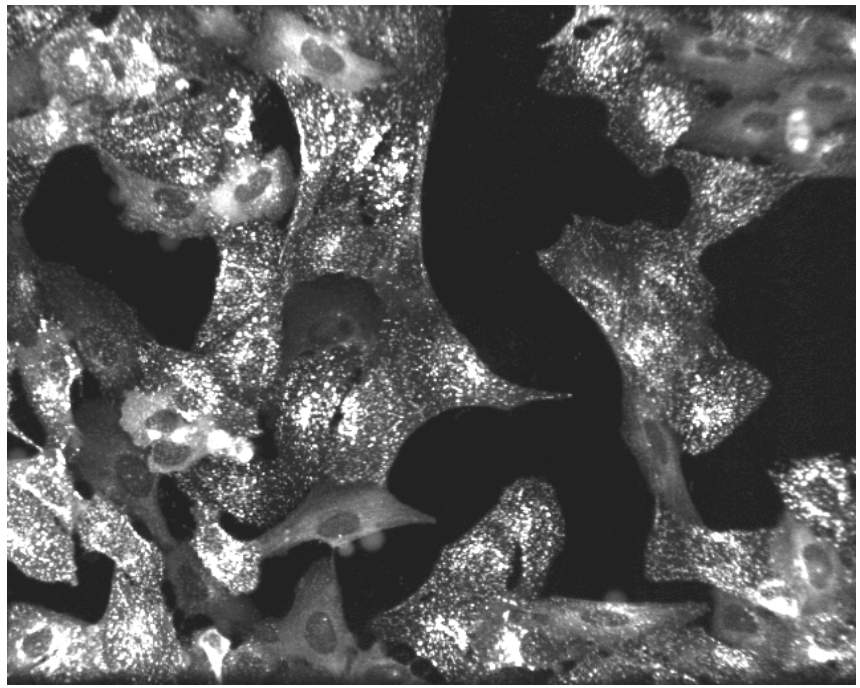


Roche Discovery Technologies

# Primary Screening Control Images

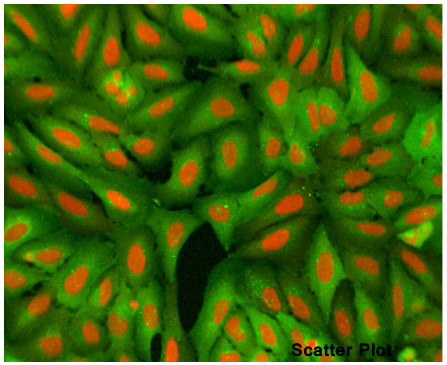


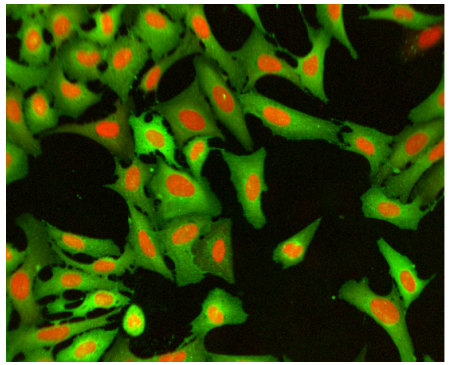
Basal

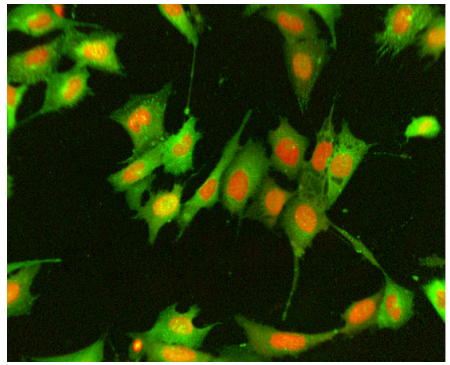


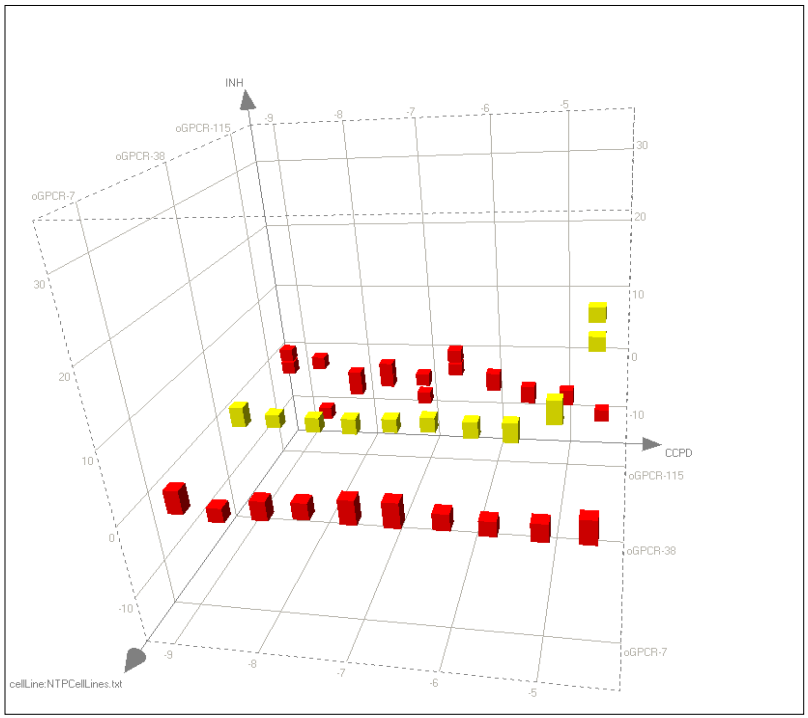
GRK (Lite)

# cmpd 27 @ 10 µM







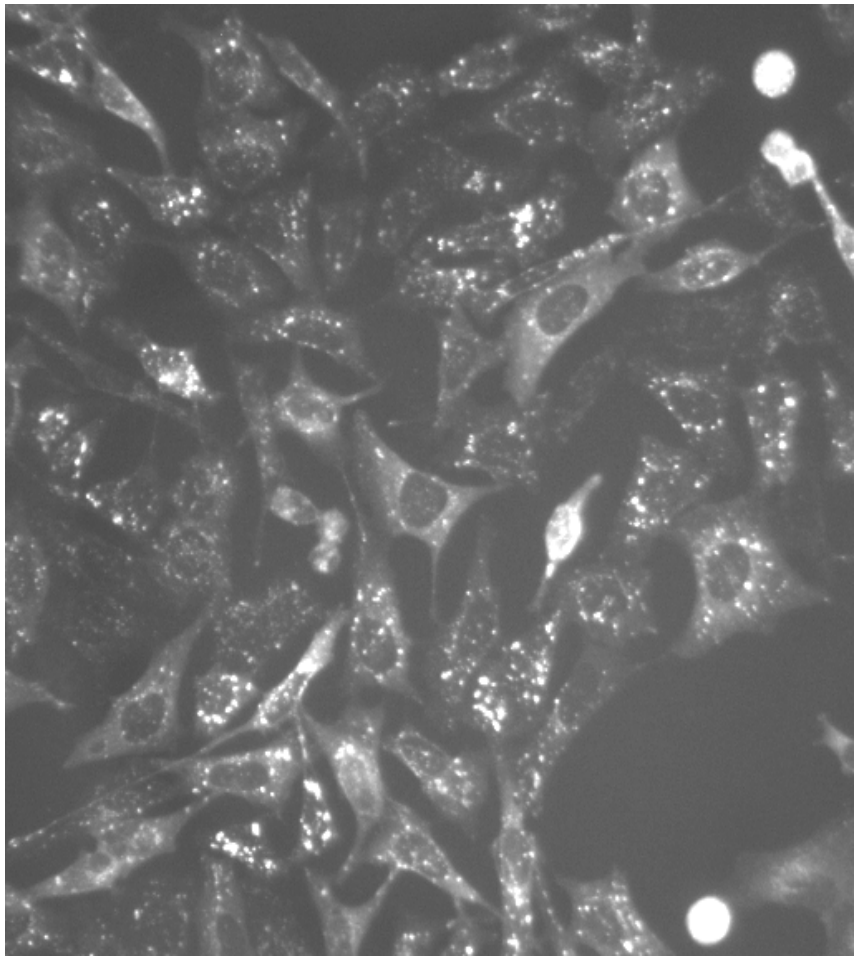

**A+ EC50-fits:**

**oGPCR-1:** yellow = exponentially diverging, Score 14

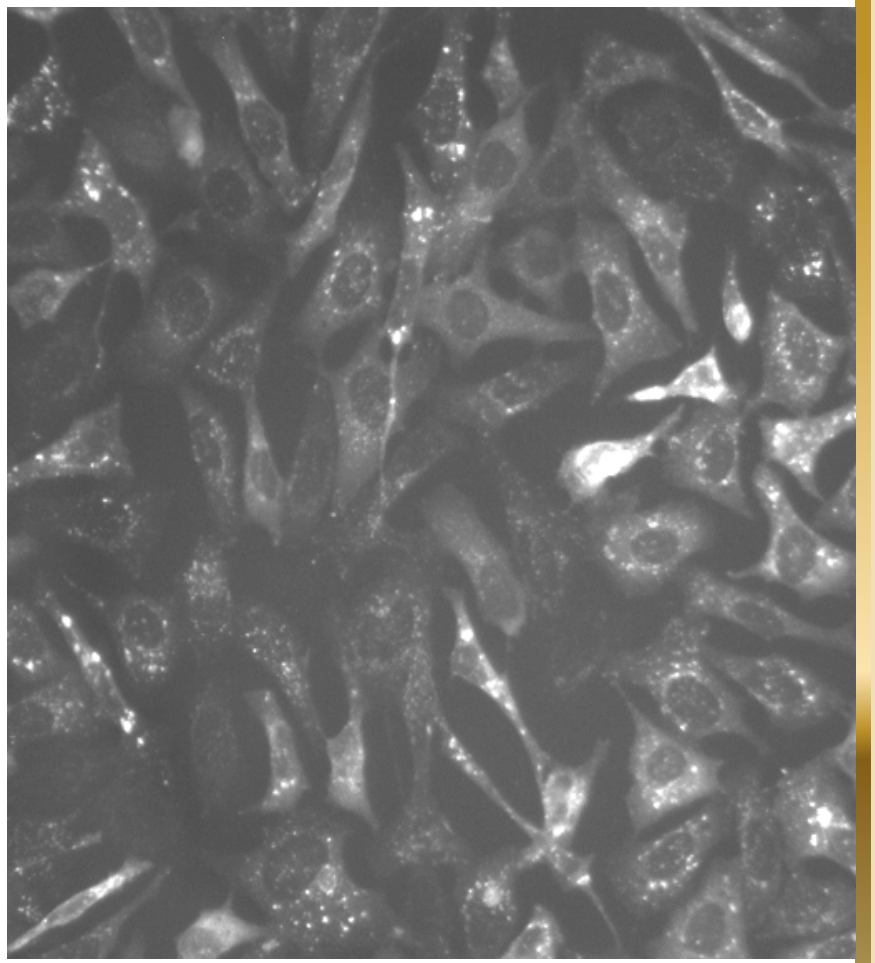
**oGPCR-2** and **oGPCR-3**: red = constant negative, Score 1

Roche Discovery Technologies

# U2OS/ oGPCR#1 cells



Well E5: 30uM RO agonist



Well E7: 10uM RO agonist

Roche Discovery Technologies