Robust Hit Identification by Quality Assurance and Multivariate Data Analysis of a High-Content, Cell-Based Assay

Oliver Dürr, François Duval, Anthony Nichols, Paul Lang, Annette Brodte, Stephan Heyse and Dominique Besson

J Biomol Screen 2007; 12; 1042
DOI: 10.1177/1087057107309036

Robust Hit Identification by Quality Assurance and Multivariate Data Analysis of a High-Content, Cell-Based Assay

OLIVER DÜRR,[1] FRANÇOIS DUVAL,[2] ANTHONY NICHOLS,[2] PAUL LANG,[2]
ANNETTE BRODTE,[1] STEPHAN HEYSE,[1] and DOMINIQUE BESSON[2]

Recent technological advances in high-content screening instrumentation have increased its ease of use and throughput, expanding the application of high-content screening to the early stages of drug discovery. However, high-content screens produce complex data sets, presenting a challenge for both extraction and interpretation of meaningful information. This shifts the high-content screening process bottleneck from the experimental to the analytical stage. In this article, the authors discuss different approaches of data analysis, using a phenotypic neurite outgrowth screen as an example. Distance measurements and hierarchical clustering methods lead to a profound understanding of different high-content screening readouts. In addition, the authors introduce a hit selection procedure based on machine learning methods and demonstrate that this method increases the hit verification rate significantly (up to a factor of 5), compared to conventional hit selection based on single readouts only. (*Journal of Biomolecular Screening* 2007:1042-1049)

Key words: phenotypic assay, high-content screening, multivariate data analysis, cellular imaging, systems cell biology, machine learning

INTRODUCTION

High-content, phenotypic screens

Modern imaging technologies together with the recent progress in cell biology pave the way to new screening paradigms.[1] The search for chemical entities modulating important cellular phenotypes, such as proliferation, differentiation, and survival, can now be conducted in biologically relevant cellular systems. In contrast to single-readout assays, high-content screening (HCS) permits the parallel measurement of an array of parameters (high content), extensively characterizing the cellular phenotype. This strategy supports the identification of small molecules based on the phenotype they elicit in a cellular environment. Active compounds such as that would be hard or impossible to find using cell-free screening approaches.

Rationale for a neurite outgrowth screen

Neurons are cells with a high capacity of regeneration. However, in pathological situations such as stroke or trauma, neurons are degenerating, leading to loss of the neuronal network. Neuroprotection thus strives for the recurring growth of neurites, which would allow this network to reassemble. Therefore, molecules inducing neurite outgrowth have a tremendous potential as neuroprotective agents.

Assessment of this "neurite outgrowth" phenotype used to be through very low throughput, nonquantitative assays. Today, new imaging technologies combined with powerful reagents and image analysis algorithms facilitate rapid and quantitative screening of a large number of compounds, increasing the chance of identifying outgrowth-stimulating compounds. However, the multitude readout parameters provided by this high-content approach require efficient data analysis environments.[2]

Multivariate data analysis on HCS screens

High-content screens produce complex data, providing a detailed molecular and phenotypic picture of the effect of compounds on cells.[3] Incorporating reference compounds, those detailed findings even permit conclusions from a compound's HCS "footprint" to the mechanistic level.[4,5]

Analysis of HCS data usually follows the image analysis steps of object recognition, image segmentation, calculation of object parameters, and calculation of population parameters.[6] Data analysis seeks 1) to ensure that the quality of the experiment and resulting data is sufficient to derive the desired conclusions, 2) to determine the readout parameters relevant to the biological question, and 3) to select compounds for further progression. Because

[1]Genedata AG, Basel, Switzerland.
[2]Merck Serono International SA, Geneva, Switzerland.

Received Apr 5, 2007, and in revised form Aug 27, 2007. Accepted for publication Aug 30, 2007.

Journal of Biomolecular Screening 12(8); 2007
DOI: 10.1177/1087057107309036

HCS is capable of providing up to 5-dimensional data sets per individual cell in some experiments,[7] data analysis can be challenging. In large-scale experiments, the population statistics of individual readout parameters (shape, intensity, contrast, localization, etc.) is usually calculated first. The resulting parameter sets (1 per well) are then subjected to further multivariate analysis, as shown in this article.

Here, cluster analysis of resulting data serves to match observed parameter correlations with their biological context and mechanisms.[5] The challenging navigation of complex HCS data sets is addressed by effective information management[8] and integrative visualization methods for cell populations and readouts.[9] Combining the information from multivariate HCS readouts increases the accuracy of decisions.[10] The method is still in its infancy, although automated parameter selection and machine learning methods such as support vector machines and other classifiers[11] are being successfully applied to this end.[12] We demonstrate such a technique for multiparametric hit selection from a high-content screen, leveraging HCS information and hereby significantly increasing precision and hit prediction.

METHODS

Experimental setup

Reagents. Neuroscreen™-1 cells and Neurite outgrowth HitKit were acquired from Cellomics (Pittsburgh, PA). RPMI-1640, penicillin/streptomycin, and L-glutamine were purchased from Gibco-Invitrogen (Carlsbad, CA). 2.5S Nerve Growth Factor (NGF) was produced by Promega (Madison, WI). Fetal bovine serum and horse serum were acquired from HyClone-Thermo Fischer Scientific (Logan, UT). Tissue culture-treated black 96-well, clear-bottom, collagen I–coated microplates were produced by BD Bioscience (San Jose, CA). Formaldehyde (37%) was acquired from Sigma (St. Louis, MO), and DMSO puriss. p.a. 99.9% was purchased from Fluka (Buchs, Switzerland).

Instrument and software. Plates were read and data analyzed using HCS reader ArrayScan® 3.5 and Extended Neurite Outgrowth Bioapplication from Cellomics (Pittsburgh, PA). The assay was set up to be supported partially by a Beckman-Coulter Biomek FX (Fullerton, CA) and partially by a Titertek MAP-C2 (Huntsville, AL).

Data produced were analyzed using the software packages Genedata Screener® (Genedata AG, Basel, Switzerland), Matlab (Mathworks, Natick, MA), and libSVM (C. C. Chang and C. J. Lin, National Taiwan University, http://www.csie.ntu.edu.tw/~cjlin/libsvm).

Neurite outgrowth assay. Neuroscreen™-1 cells were cultured in RPMI-1640 supplemented with 2 mM L-glutamine, 10% horse serum, 5% fetal bovine serum, and 100 units of penicillin/streptomycin. Collagen I–coated 96-well microplates were prepared on a Biomek FX; the following 50-µL controls were prepared using medium with 0.2% DMSO (negative control: wells A1 to D1), 0.6 ng/mL NGF (EC_{25}: wells E1 to H1), 2 ng/mL NGF (EC_{50}: wells E12 to H12), and 200 ng/mL NGF (positive control: wells E12 to H12); all the other wells contained 50-µL compounds diluted in medium to a final concentration of 20 µM and 0.2% DMSO. A cell suspension was diluted with culture medium at 5×10^4 cells/mL, and 50 µL of cell suspension was seeded into the microplates. The cells were incubated 3 days at 37 °C in a humidified atmosphere of 5% CO_2. The next steps were performed on a MAP-C2 at room temperature. The medium was removed, and 100 µL phosphate-buffered saline (PBS) 3.7% formaldehyde and 1:2000 diluted Hoechst Dye Solution were added to each well. After a 20-min incubation, the fixation solution was removed, and the plates were washed 3 times with 100 µL 1× Neurite Outgrowth Buffer. The Neurite Outgrowth Buffer was removed, and 50 µL of Primary Antibody solution was added to each well. After a 1-h incubation, the Primary Antibody solution was removed, and the plates were washed 3 times with 100 µL 1× Neurite Outgrowth Buffer. The Neurite Outgrowth Buffer was removed, and 50 µL of Secondary Antibody solution was added to each well. After a 1-h incubation, the Secondary Antibody solution was removed, and the plates were washed 3 times with 100 µL 1× Neurite Outgrowth Buffer and twice with 100 µL PBS. The wells were kept in 200 µL of PBS; the plates were sealed and read on an ArrayScan HCS Reader using a 10× objective and the Extended Neurite Outgrowth Bioapplication using 2 channels (channel 1: Hoechst nuclear dye; channel 2: total neuron dye).

In total, 40 parameters were extracted from the images. A representative selection is given in **Table 1**.

Data analysis methods

Relationship between HCS readouts. A common measure to investigate the relationship between 2 readouts, x_1 and x_2, is Pearson's correlation coefficient *r:*

$$r = \frac{\sum z(x_1)\, z(x_2)}{n-1}. \qquad (1)$$

It is calculated from the *z*-transformed readouts, x_1 and x_2, for each well in a screen containing *n* wells. The *z*-transform is obtained by subtracting the mean from each individual measurement x_i and then dividing by the standard deviation. The correlation coefficient ranges from +1 (perfect correlation) to −1 (perfect anticorrelation). A value of 0 indicates that there are no (linear) correlations between the 2 readouts. Typically, *r* is calculated between each pair of readouts, resulting in a (symmetrical) matrix. To facilitate the interpretation of this matrix, a hierarchical cluster tree or dendrogram can be derived using Ward's clustering method.[13]

Classification. In a workflow combining multiple HCS readouts for hit selection, automated classification is a key method for

Table 1. Details of Read-Out Parameters (Selection)

Read-Out Parameter	*Definition*
MEAN_TotalNeuriteLengthCh2	The well average of total neurite length for all neurons detected using channel 2
%OutgrowthPositiveCh2	Percentage of neurons detected using channel 2 in the well that have positive outgrowth of neurites
%TotalNeuriteLengthPositiveCh2	Percentage of neurons detected using channel 2 in the well that have total neurite length greater than the total neurite length threshold
MEAN_AvgNeuriteLengthCh2	The well average of the average neurite length for all neurons detected using channel 2
MEAN_NeuriteCountCh2	The well average of neurite count for all neurons detected using channel 2
ValidNucleusCount	Count of valid nuclei in the well
%ValidNeuronsPerNucleus	Percentage of valid cells in the well that are valid neurons
MEAN_CellBodyAvgIntenCh2	The well average of cell body average intensity for all neurons detected using channel 2
SE_NeuriteCountCh2	The well standard error of neurite count for all neurons detected using channel 2
SE_AvgNeuriteLengthCh2	The well standard error of average neurite length for all neurons detected using channel 2

the robust prediction of potent compounds. Therefore, we give a short introduction to classification; more comprehensive reviews can be found in the literature.[14]

Classification is performed in 2 phases. First, in the training phase, a set of known (reference) compounds each forms set S_1 or S_2, respectively. Their readouts are used to train a classifier. Set S_1 could contain active reference compounds, whereas set S_2 could consist of inactive compounds. Second, in the classification phase, the classifier is used to assign the screened compounds to either class S_1 (actives) or to class S_2 (inactives), based on the compound's readout profiles.

There is an array of classification algorithms from which to choose. Most of them have parameters, which can be tuned. Before applying any classification to new data, both algorithms and parameters should be systematically evaluated, using a cross-validation scheme. In this scheme, the set of known reference compounds is randomly and repeatedly divided into 2 groups: 1 training set and 1 test set. Then a classification is performed, recording how many compounds are being classified correctly. The parameters then can be fine-tuned to increase the percentage of correctly classified compounds (i.e., the accuracy of classification).

In this article, we use the following classification algorithms: K-nearest neighbors (K-NN), Fisher linear discriminant analysis (LDA), and support vector machines (SVM) with linear and with Gaussian kernel.

All these algorithms work in an n-dimensional space spanned by the n readouts. A single measurement is represented in that space by a vector whose components are the measured values for the individual readouts.

Normalization was carried out using the positive and negative controls for the activity readouts, whereas no normalization was performed for the nonactivity readouts. After normalization, all readouts were z-transformed to scale them to one and the same range.

In Fisher LDA, a linear combination of all readouts forms a vector $\vec{w}$, optimally separating the sets S_1 and S_2.[15] The unknown measurement $\vec{x}$ is classified to belong to S_1 if the scalar product of $\vec{x}\vec{w}$ is positive. Otherwise, it is assigned to S_2. In the K-NN classification, the K-nearest neighbors of $\vec{x}$ are searched, and $\vec{x}$ is assigned to the same class to which the majority of its neighbors belong. Common distance metrics to determine the nearest neighbors are Euclidian or correlation-based distances; we have used Euclidean distance.

For the SVM classification, we applied a generalization of the simple linear SVM in 2 ways. In method 1, we used a soft margin, which makes the SVM applicable to the nonseparating case.[16] In this case, the location of S_i elements in the "wrong" part of the subspace is permitted but penalized. The total deviations are penalized proportionally to a user-definable parameter C that determines the softness of the margin: large values correspond to a rigid margin (high penalty of misclassified items).[16]

The second method is a transformation to a higher dimensional space, prior to classification.[17] Using the "kernel trick," this transformation does not need to be carried out explicitly if a nonlinear kernel $k(\vec{x},\vec{x}')$ replaces every dot product in the original space $\vec{x}\vec{x}'$. In our study, we used the widely established Gaussian kernel, defined via

$$k(\vec{x},\vec{x}') = e^{-\gamma\|\vec{x}-\vec{x}'\|^2}. \quad (2)$$

To rank the compounds in linear SVM, we used the distance to the hyperplane. For nonlinear kernels, this distance can be generalized to the affinity. The affinity $A(\vec{x})$ of a compound with the readout vector $\vec{x}$ is defined as

$$A(\vec{x}) = \sum_i Y_i \alpha_i K(\vec{V}_i, \vec{x}) + b, \quad (3)$$

where the sum is taken over all support vectors $\vec{V}_i$; Y_i is –1 if V_i is part of S_1 and 1 otherwise, b is the distance of the hyperplane to the origin, and α_i are the Lagrange multipliers of the optimization problem. For the linear case, the kernel $k(.,.)$ simplifies to the standard scalar product, and the equation above is proportional to the Euclidian distance to the hyperplane.[14]

Hit selection procedures. In high-throughput screening, hits are traditionally picked by ranking all compounds by 1 single-activity

readout value first and before applying a threshold. This threshold is determined either by data distribution alone (e.g., 3 standard deviations from the mean of all compound or control measurements) or based on the capacity available for rescreening (e.g., the 1000 most active compounds). As this "conventional" approach only takes 1 readout into account, its application to high-content screens neglects the information provided by additional readouts. To improve this process, additional readouts can be used as filters. Wells where the respective readouts do not fall within a certain range (e.g., images with an insufficient number of cells present) will then be rejected. This improves the quality of the hit list by eliminating wells with clearly nonreliable measurements, reducing false positives.

Another way to exploit the information contained in multiple readouts is to apply the classification methods described earlier. A classifier is trained using the reference compounds and subsequently applied to predict the similarity of a screened compound to the active reference compounds, using all readouts. The affinity *A* is then used to rank the compounds according to their expected activity. For experiments done in replicates, the affinity is calculated for all individual measurements. For compound ranking, the median affinities of the individual measurements are used.

To eliminate potential errors in some measurements of the reference compounds, a cross-validation of wells containing reference compounds is performed before classification. Wells containing reference compounds not predicted into their own class are removed from the training set.

RESULTS AND DISCUSSION

A set of 7000 compounds was screened in duplicate on 96-well plates. Each plate had active reference controls at 3 different concentrations (EC_{25}, EC_{50}, and positive control) and negative control (example images in **Fig. 1**). All controls were applied in quadruplicate. The remaining 80 wells per plate contained compounds.

The first quality control (QC) step was performed using a filtering approach by QC parameters. Six parameters were chosen by biological reasoning to filter out low-quality measurements, and their respective thresholds were set by experience (**Table 2**).

After quality filtering, overall assay robustness was good as it fulfilled defined global quality criteria such as a robust Z′ (RZ′) above 0.5 and a low percentage of discarded wells (6%), including 3 full plates. These wells were excluded in all subsequent analyses. The hit selection has subsequently been based on a single parameter (MEAN_AVG_NEURITE_LENGTH) that has shown to be the most robust and relevant during the assay development phase.

Correlation between individual HCS readouts

To investigate the statistical correlation between the readouts, a correlation analysis and hierarchical clustering were performed. The results are displayed in **Figure 2**. The correlation analysis shows a strong correlation between, for example, the readouts in the upper left corner (%TotalNeuriteLengthPositive, MEAN_AvgNeuriteLength, OutgrowthPositiveCount, %OutgrowthPositive, MEAN_NeuriteCount, MEAN_TotalNeuriteLength), all having to do with neurite outgrowth. Readout parameters colored in yellow such as "OutgrowthPositiveCount" are considered relevant for

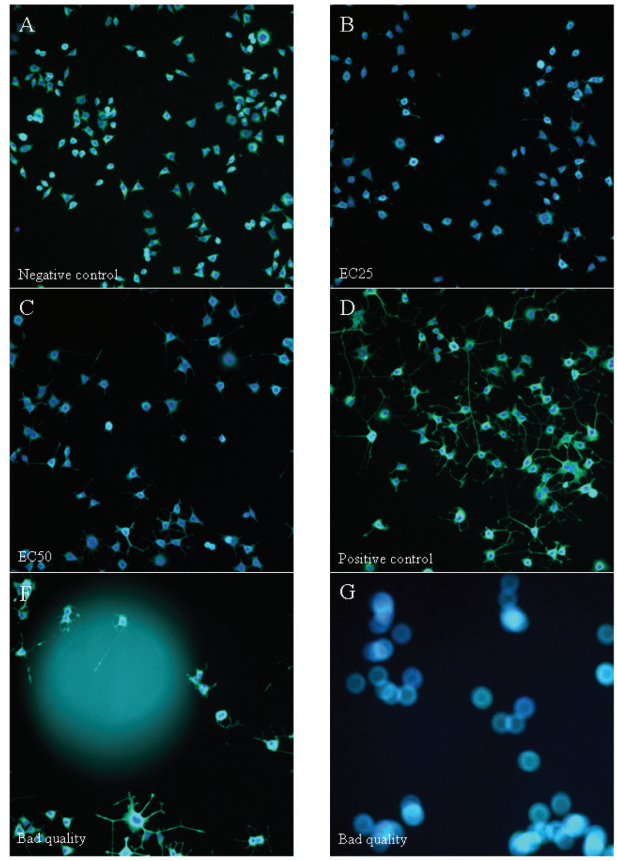

FIG. 1. Example images from the neurite outgrowth screen. Displayed are the Neuroscreen™-1 cells with the nuclei shown in blue and the cytoplasm in green. (**A**) Negative control, (**B**) NGF-treated cells of EC_{25} concentration, (**C**) EC_{50} NGF-treated cells of EC_{50} concentration, (**D**) positive control cells, and (**E, F**) bad-quality images rejected during the quality control process.

Table 2. Parameters Used for Automated Quality Filtering

Parameter Name	*Defined Limit*	*Defined Value*
ValideNucleusCount		80 to 400 cells
%ValidNeuronsPer Nucleus	2 SD	42%
MEAN_CellBody AvgIntenCh2	2 SD	221.6
MEAN_Neurite CountCh2	2 SD	6.9
SE_NeuriteCountCh2	2 SD	0.4
SE_AvgNeurite LengthCh2	3 SD	22

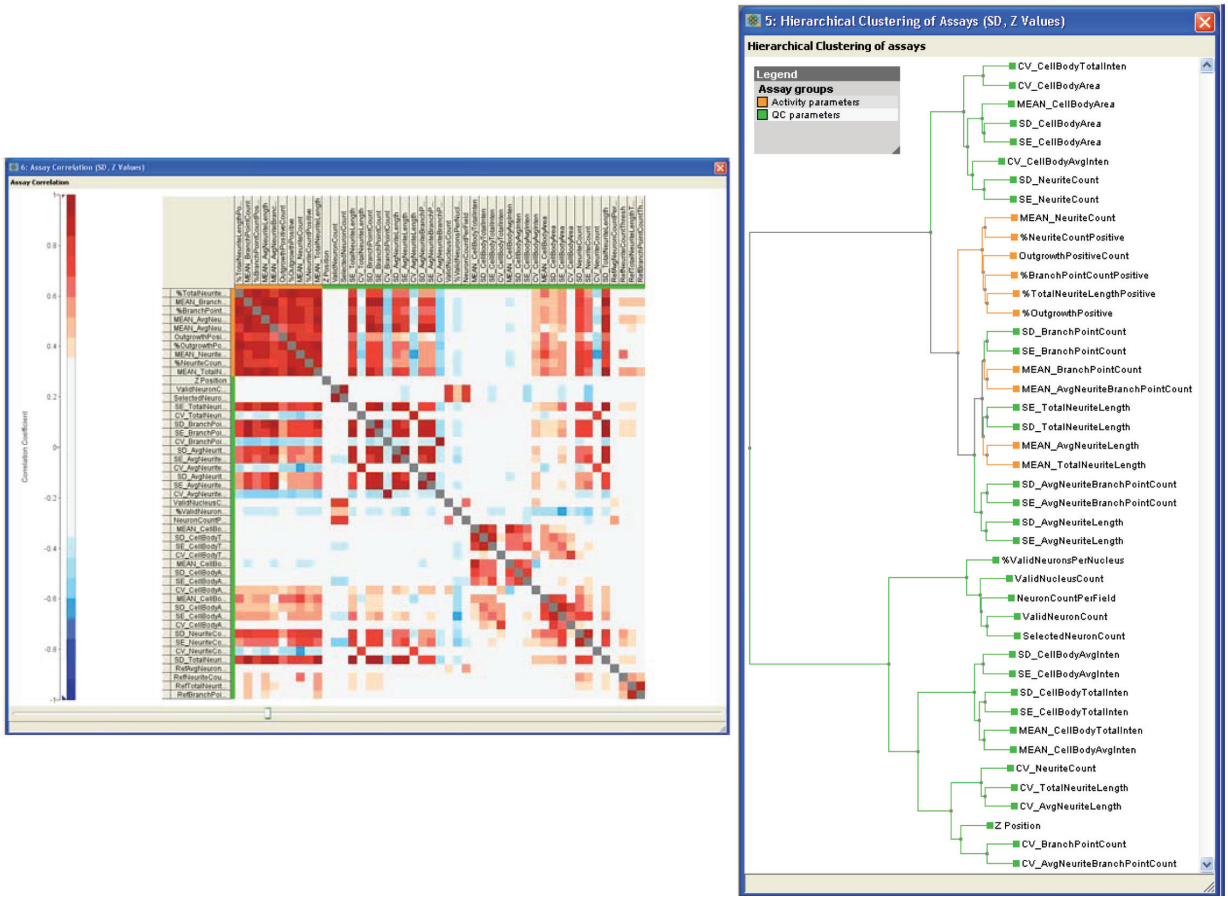

FIG. 2. Matrix of all correlation coefficients between the readout parameters (left) and their hierarchical clustering based on this correlation (right). Readout parameters colored in yellow are considered relevant for activity due to their biological meaning. Screenshot taken from Screener® Sarileo.

activity due to their biological meaning. Using the same data, but displaying them via a cluster tree, the hierarchical data structure is revealed. The cluster tree shows that the readouts cluster into 5 major classes. The first class from the top contains the cell body area, as well as its errors and the errors of the NeuriteCount, indicating an influence of cell size on the neurite counts error. The second class contains mainly activity readouts. The third class is a mixture of the branch point counts and errors of the neurite length, showing that the neurite length errors are correlated to the branch points. The remaining 2 classes are somewhat separated from the rest and contain the remaining readouts.

An automated hit-picking procedure

In this section, the results of a standard threshold-based hit-picking procedure and an advanced automatic workflow are compared. To evaluate the performance of the respective procedures, 156 compounds have been rescreened in an additional experimental run. This set of 156 compounds was composed of 84 compounds suggested by the machine learning procedure and 85 by the thresholding procedure, with an overlap of 13 compounds.

The readouts have been restricted by biological reasoning to the following set of 8 readouts: %ValidNeuronsPerNucleus, SelectedNeuronCount, MEAN_CellBodyAvgIntenCh2, MEAN_NeuriteCountCh2, SE_NeuriteCountCh2, MEAN_AvgNeuriteLengthCh2, SE_AvgNeuriteLengthCh2, and %OutgrowthPositiveCh2. To restrict the set of readouts, we compared the performance of a cross-validation on the set of 8 readouts against the set of all readouts, using the EC_{25} and negative control.

Using the SVM classifier ($\gamma = 0.02$, $C = 10$), the restricted set of 8 readouts had a correct classification rate of 98%, whereas the set of all readouts had only 96%. Because including more readouts seemed not to increase performance significantly, the set of 8 readouts was used subsequently. Furthermore, due to the generally high verification rates in cross-validation, using the controls,

Table 3. Cross-Validation Using the Primary Data on the Selected 8 Readouts

Algorithm	*Wells Correctly Predicted to EC_{25} Controls, %*	*Wells Correctly Predicted to Negative Controls, %*
K = 1 NN	95.8	96.2
K = 2 NN	94.3	98.8
LDA	93.4	98.5
SVM (Gaussian-kernel) $\gamma = 0.02$, $C = 10$	98.3	98.4

K-NN, K-nearest neighbors; LDA, linear discriminant analysis; SVM, support vector machine.

a systematic comparison of the predictivity of different subsets of parameters would not have been very significant.

For the standard approach, the MEAN_AVG_NEURITE_LENGTH parameter with a threshold of 3 standard deviations (20% normalized signal) was used.

The classifier was trained using the EC_{25} and the negative control. To find the best classifier, we applied a 100-fold cross-validation with a test set fraction of EC_{25}. The results are displayed in **Table 3**.

The SVM using a Gaussian kernel with our standard settings ($C = 10$, $\gamma = 0.02$) and $k = 2$ NN showed superior results compared to $k = 1$ NN and Fisher LDA classifications. As an additional advantage over K-NN, SVM provides a continuous ranged affinity to the reference compounds, allowing for fine-tuned control over the entire hit list. SVM was therefore used for activity prediction of the screened compounds as follows: because the correct classification rate was already very high, it seemed not reasonable to optimize SVM parameters any further. A sensitivity analysis demonstrated that the parameter settings used for γ and C were close to optimal.

The 84 best compounds using the machine learning approach and the 85 compounds with the highest values in the readout "MEAN_AVG_NEURITE_LENGTH" ("traditional approach") underwent elaborate retesting, ensuring that data from the rescreen were a more reliable estimate ($RZ' = 0.63$) of the compounds' true activity.

When using a cutoff of 3 standard deviations (negative controls) of the parameter "MEAN_AVG_NEURITE_LENGTH" in the rescreen, we found 48 hits; 44 (91.7%) of them had been suggested by the SVM approach, whereas only 10 (20.8%) had been proposed by the "traditional" approach.

Hence, among the compounds with the highest activity in the rescreen, the machine learning approach did suggest 4 of 5 compounds correctly. **Figure 3** shows that this superior performance does not significantly depend on the chosen cutoff. For a reasonable number of hits, SVM is superior to the traditional method by a factor of up to 5.

A posteriori analysis of the classification parameters

The 48 compounds with the highest activity in the retest were labeled as hits; the rest were labeled as nonhits. The data from the primary screen were used to perform a cross-validation study. The 2 parameters (γ and C) of the SVM were varied and the performance recorded. The result is displayed in **Figure 4**. For the analysis described above, we used $C = 10$ and $\gamma = 0.02$, which has a verification rate of 74%, close to the optimum of 74.5% at $C = 10$ and $\gamma = 0.04$.

CONCLUSIONS

The work presented here shows detailed ways for the analysis of cell-based high-content screens, managing their inherent complexity and leveraging their information content for robust hit identification. The assay used in this study measured the stimulation of neurite outgrowth by small molecules, a phenotype of high pharmacological potential. Screens of this type are not amenable to traditional screening procedures but require automated microscopy for detection of active compounds. In such a complex assay setup, a robust assay protocol with low variation and high reliability is key. For this assay, the robust Z prime factor of $RZ' > 0.5$, and the relatively low fraction of 6% of low-quality wells masked by the filtering procedure described earlier indicates assay robustness.

HCS data are multivariate by their very nature; cellular phenotypes are simply too complex to be described by one variable only. Image analysis software thus extracts several readouts for each cell from the microscopy images, preserving information. However, traditional hit-ranking approaches are based on a single (measured or calculated) readout for compound ranking. We showed that this approach is inadequate when applied to high-content screens and demonstrated alternative techniques to analyze in parallel and to leverage all readouts.

Data analysis techniques such as the correlation matrix and hierarchical clustering allowed us to determine the relationship between parameters and to compare these findings with expectations based on the assay's biology. In our study, the evaluation of readout parameters related well to assay biology, indicating that measurements were sensible and image analysis worked perfectly. Thus, it represents an important QC technique for both measurement and image processing.

Finally, we have introduced a concept to classify compounds using machine learning techniques according to their ability to produce the "activated" cellular phenotype. We validated this concept by experimentally rescreening the predicted actives. We have found that the machine learning approach had a 5 times better prediction rate, compared to traditional hit list generation. The main reason for this outcome is that multivariate classification makes better use of the wealth of information

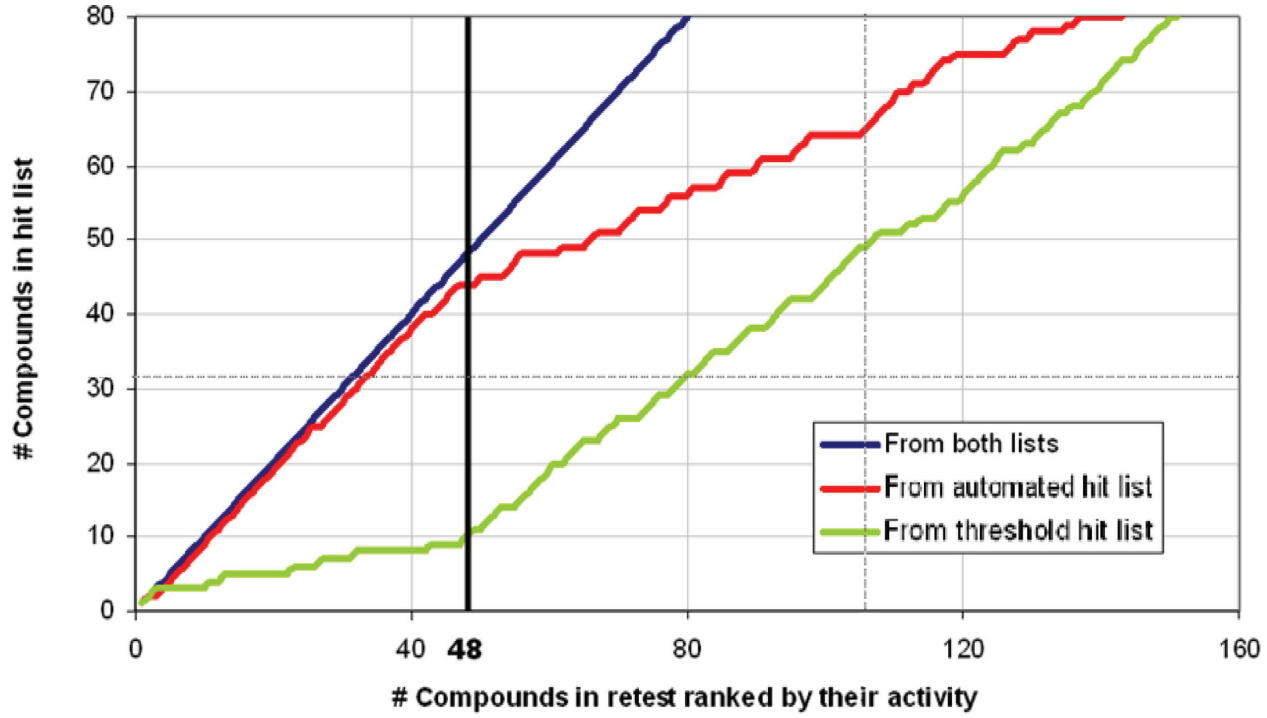

FIG. 3. Top scorers in the retest originating either from the automated hit-picking procedure (red) or the traditional approach (green). On the *x*-axis, the retested compounds are ranked according to their activity in the retest. The vertical bar corresponds to a cutoff of 3 SD (negative controls) of the parameter "MEAN_AVG_NEURITE_LENGTH," a usual cutoff taken for the hit list.

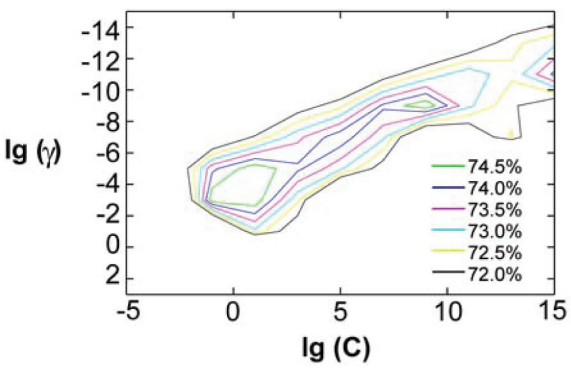

FIG. 4. Cross-validation rate using a support vector machine with Gaussian kernel as a function of the kernel parameter γ and the softness of the margin *C*. The values used for the hit prediction are lg(γ) = –5.64 (γ = 0.02) and lg(*C*) = 3.32 and correspond to a verification rate of 74%.

from HCS images, whereas the traditional approach relies on a single parameter only.

In summary, we have demonstrated how multivariate statistical techniques allow easy handling and leveraging of the complex information provided by a high-content assay. Using these techniques, the final results obtained from such a screen are of higher precision and much more informative. They justify the investments in a more expensive measurement technology, reduce the number of follow-up experiments, and increase the success rate in identifying lead candidates.

ACKNOWLEDGMENTS

The authors thank Yves Sagot and Astrid Osen-Sand for their scientific input, Olivier Bitterlin for the essential support on the compound management side, and Christèle Fremaux, Christophe Cleva, and Ioannis Xenarios for the help in managing and analyzing screening data. Oliver Leven is acknowledged for the fruitful discussions, are as Daniel Domine, Christiane Becker, and James Kristie for their careful proofreading of the manuscript.

REFERENCES

1. Lang P, Yeow K, Nichols A, Scheer A: Cellular imaging in drug discovery. *Nat Rev Drug Disc* 2006;5:343-356.
2. Taylor DL, Haskins RH, Giuliano KA: *High Content Screening*. Totowa, NJ: Humana Press, 2007.

3. Haney SA, LaPan P, Pan J, Zhang J: High-content screening moves to the front of the line. *Drug Disc Today* 2006;11:889-894.
4. Perlman ZE, Slack MD, Feng Y, Mitchison TR, Wu LF, Altschuler SJ: Multidimensional drug profiling by automated microscopy. *Science* 2004; 306:1194-1198.
5. Guiliano KA, DeBiasio RL, Dunlay RT, Gough A, Volosky JM, Zock J, et al: High-content screening with siRNA optimizes a cell biological approach to drug discovery. *J Biomol Screen* 2004;9:557-568.
6. Abraham VC, Taylor DL, Haskins JR: High-content screening applied to large-scale cell biology. *Trends Biotech* 2004;22:15-22.
7. Swedlow JR, Goldberg J, Brauner E, Sorger PK: Informatics and quantitative analysis in biological imaging. *Science* 2003;300:100-102.
8. Dunlay RT, Czekalski WJ, Collins MA: Overview of informatics for high content screening. *Methods Mol Biol* 2007;356:269-280.
9. Giuliano KA, Cheung WS, Curran DP, Day BW, Kassick AJ, Lazo JS, et al: Systems cell biology knowledge created from high content screening. *Assay Drug Dev Tech* 2005;3:501-514.
10. Gasparri F, Mariani M, Sola F, Galvani A: Quantification of the proliferation index of human dermal fibroblast cultures with the ArrayScan high-content screening reader. *J Biomol Screen* 2004;9:232-243.
11. Bhaskar H, Hoyle DC, Singh S: Machine learning in bioinformatics: a brief survey and recommendations for practitioners. *Comput Biol Med* 2006;36:1104-1125.
12. Heyse S, Brodte A, Bruttger O, Dürr O, Freeman T, Jung T, et al: Quantifying bioactivity on a large scale: quality assurance and analysis of multiparametric ultra-HTS data. *J Assoc Lab Automation* 2005;10:207-212.
13. Ward JH: Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 1963;58:235-244.
14. Duda RO, Hart PE, Stork DG: *Pattern Classification*. 2nd ed. Indianapolis: Wiley-Interscience, 2000.
15. Fisher RA: The use of multiple measures in taxonomic problems. *Ann Eugenics* 1936;7:179-188.
16. Cortes C, Vapnik V: Support-vector networks. *Machine Learning* 1995; 20:273-297.
17. Boser BE, Guyon IM, Vapnik V: A training algorithm for optimal margin classifiers. In Haussler D (ed): *5th Annual ACM Workshop on COLT*. Pittsburgh: ACM Press, 1992:144-152.

Address correspondence to:
Oliver Dürr
Genedata AG
Maulbeerstrasse 46
4016 Basel, Switzerland

E-mail: oliver.duerr@genedata.com