WHITE PAPER

Improving the Drug Discovery Tool Box:
How to Achieve More Efficient Processing and Analysis of Screening Data

The data deluge. The elusive molecule. Activating the receptor. The greater complexity of protein-based therapeutics. Rapidly identifying side effects. The financial impact of success or failure.

These are only a few considerations in the lengthy series of choices required to select a drug target. The staggering amount of screening data that must be processed in choosing candidates is well known to the industry. Meanwhile the technologies applied to the winnowing of lead candidates emerging from High Throughput Screening (HTS) continue to increase in complexity.

Decisions matter. Making good decisions requires software to effectively manage the size and complexity of multiparametric datasets. Companies that most quickly and accurately identify key drug discovery parameters will win the race to identify the most appropriate targets and develop medicines to affect them.

Scientists are tasked with finding ways to screen for potential problems with promising molecules at the earliest possible stage. Achieving this requires an accessible and intuitive screening solution to analyze and review data from multiple outputs within a single platform. By facilitating the comparison of data from different assay types, scientists then have confidence that those assays are giving accurate results, leading to substantial improvements in the drug discovery process, and ultimately health outcomes.

High-Throughput Screening Data

The best-known and longest-running approach to accelerating drug discovery is high throughput screening (HTS), conducting assays in high-density multi-well microtiter plates with data coming from microplate readers. The goal: the production of large amounts of data in flat files. Promising compounds, a small number within the overall quantity, are then tested further to determine if they are worthy of the time and expense of additional evaluation. This process can result in workflow bottlenecks, because the level of detail required for analysis is immense.

High-Content Screening Data

Unlike HTS, High Content Screening (HCS) data is generated by various high-resolution fluorescent or confocal microscopes rather than by plate readers. While the overall throughput in HCS is not as great as HTS, the information about the impact of the candidate compound can lead to a better understanding of its effect on cells.

However, this comes at a cost in terms of data management and analysis. The amount of information generated per well in HCS is much higher than in HTS, presenting quantities of raw data per plate that are some of the highest in biopharma. The complexity of the information is also higher, creating additional challenges (and bottlenecks) in data processing and analysis.

PerkinElmer®
For the Better

Surface Plasmon Resonance

Surface Plasmon Resonance (SPR) offers substantial benefits for binding measurements in biochemical and pharmacological studies. Measurements are sensitive and require small amounts of reagents. SPR is differentiated by its ability to generate both equilibrium binding data and measurements of the kinetics of interactions. For a promising lead identified by HTS or HCS to progress to drug candidate status, it is usually necessary to characterize its binding affinity and the association and dissociation kinetic rate constants involved in its binding to the target. Additionally, with the increasing importance of macromolecules as therapeutic agents, particularly monoclonal antibodies (mAbs), characterization of the epitopes of the mAb binding to the target is sometimes desired too.

SPR has become the gold standard for obtaining these measurements. SPR has additional advantages over other techniques. For example, SPR does not require a label for detection (molecular labels can impact the binding of compound to the target), and SPR requires very little target material (which can be expensive to generate/purify with protein molecules). The cost of these advantages is that truly high throughput, i.e., to the levels seen in HTS and even HCS is not possible.

Typically, the analysis software for reading SPR instrument data is inflexible, demanding considerable work to get the data into a format appropriate for reporting/data storage. The manipulation of data for screens of even a few plates can require several days to analyze and format. When many different SPR instruments are present in the research environment, each with its own analysis software, users are further burdened by the need to learn multiple packages.

Bigger Data. Better Workflow

Applying the methodologies cited above, scientists hope to screen more samples in less time, with less labor, while picking up on problems with promising compounds early in the discovery process. Achieving this comes from faster access to big data and the ability to integrate that data into an efficient workflow.

PerkinElmer Informatics has developed applications integrated within the powerful TIBCO Spotfire® data analytics platform that support the data volumes customers require. Signals Screening is a flexible and intuitive data processing engine embedded seamlessly within the familiar TIBCO Spotfire® interface. It is designed to relieve restrictions in data analysis workflows across the biopharma industry, accelerating drug discovery.

Signals Screening processes raw data from multiple instrument sources, both those from PerkinElmer as well as other manufacturers. Researchers do not need to learn multiple software packages for the import and processing of instrument data. The underlying data tables within TIBCO Spotfire® help increase data integrity by eliminating the need for cutting and pasting between external programs such as Excel. TIBCO Spotfire® provides advanced statistical analysis such as multiparametric data processing algorithms, not commonly provided, increasing the data processing power available to drug discovery scientists, offering deeper insights into the data.

Signals Screening's intuitive interface makes it easily configured for rapidly processing large datasets from a variety of common drug discovery applications, including the output of plate readers running basic HTS assays, image processed HCS data for phenotypic, multiparametric analyses, and kinetics data from SPR instruments.

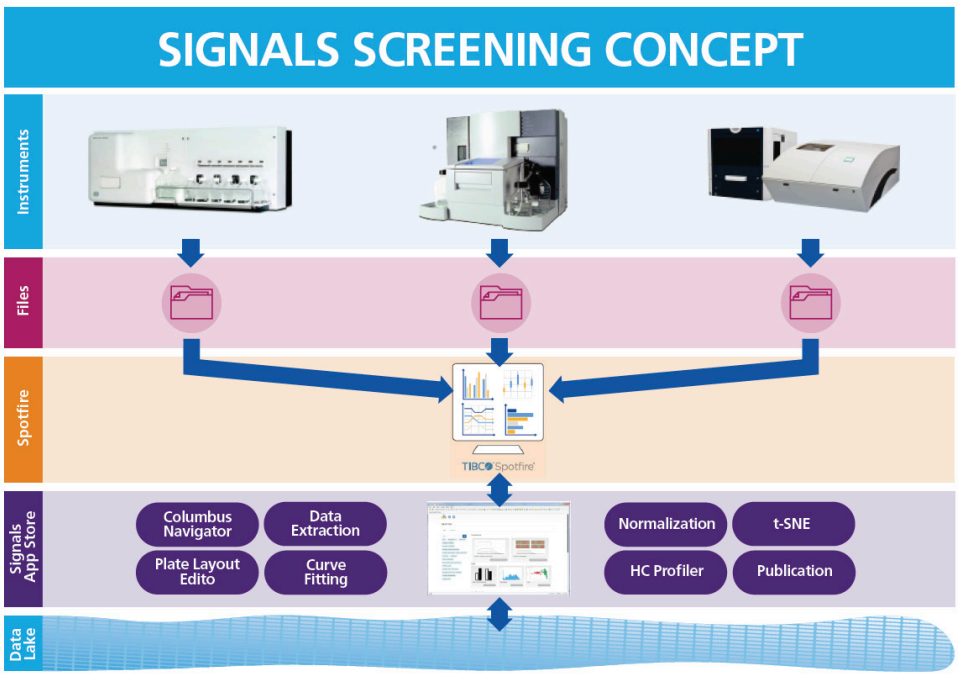

Figure 1. High Content Screening (HCS) combines automated confocal microscopy and quantitative image analysis for acquisition of multi-parametric data from single cells.

Additional productivity gains are enabled through the use of Signals Screening protocols applied to procedures that run regularly. These offer near-automation levels for importing data, formatting plates with metadata, normalizing data relative to controls, and curve fitting. In addition to saving time, Signals Screening helps reduce opportunities for human error and increases the analysis consistency from individual scientists, within groups, and across the enterprise. Integration of Signals Screening output with electronic data notebooks, or by publishing to corporate databases, is also available and reinforces data integrity.

The initial offering of Signals Screening addresses three key drug discovery functions: HTS, HCS, and SPR.

Signals Screening for Basic Screening Applications

The key concept of Signals Screening-Basic Screening is that various instrument files are stepped through user-configured, fit-for-purpose workflows for processing in TIBCO Spotfire®. In the design of Signals Screening, PerkinElmer has componentized various elements of commonly used screening operations, then packaged these into an integrated data analysis tool that leverages TIBCO Spotfire®'s industry-leading analytics and data visualization capabilities. Once TIBCO Spotfire® is opened, the Signals Screening application is presented as one of the options. Specific workflows are built up from granular units referred to as apps; these apps are selected from the Signals Screening apps interface.

The structure of Signals Screening-Basic Screening workflows adheres to logical and straightforward procedures that scientists follow in managing screening data; plate reader data is first imported, then the meaning of each of the wells on the plate is defined based on a user-specified plate format. Next the data is normalized, then regression analysis is applied by the user's choice of curve-fitting algorithms.

Data Import: The user's ability to readily manage multiple instrument data formats is one of Signals Screening's key advantages. PerkinElmer has built into Signals Screening an exceedingly flexible and fully featured data parser to characterize plate features – i.e., the types of measurements taken. After selecting the Data Import app users select the format used to generate the data. Data is parsed appropriately based on the metadata characteristic of the selected data format and files are obtained.

Plate Editor: The Plate Editor is the app where plate maps are designed or reused. In designing the map, users apply annotations to the plate – the well format, where the positive and negative controls are located, the concentrations of test articles and controls in wells, and any other details necessary to capture for a given experimental design. If the plate layout has been previously created, whether as a shared or local design, it can be selected from the library to accept maps already applicable to the work at hand or to add new annotations and values.

Normalization: Normalization is generally the next step in data processing. Signals Screening offers two options for normalizing data, one for standard techniques and another for custom, user-specified normalizations. The Normalization app in Signals Screening provides several standard normalization methods in a pulldown menu, currently including normalization by percent inhibition, by plate median, or by use of standard paralog ratio test (PRT). With any of these methods, users simply identify which columns designate the positive and negative controls. Once a normalization method has been selected and applied, graphical representations of the data before and after normalization are displayed side-by-side for ready comparison by users.

Custom Calculation: For users who require a customized solution for normalizing their data, Signals Screening includes the Custom Calculation app.

Curve Fitting: Once the data has been normalized, users are ready to apply various analytical regression methods in fitting the data to a curve. Algorithms that work best for data are selected from a pulldown list after quickly assessing the performance of different curves within a given regression. Upper and lower asymptotes can be easily adjusted to correct for curve fit errors or address specific needs. Outlier data points are also readily identified and removed to improve the quality of the curves. Modified and unmodified curves are visualized side by side for immediate comparison and realization of data quality control. Selected curves are saved for publication to the data lake.

Using the powerful arsenal of Signals Screening apps, users can establish a data processing workflow all the way from data import to data QC in a single session. The exact steps of that workflow can be saved as a protocol for easy reuse in future sessions. Because Signals Screening captures all relevant metadata from previous analysis, it is easy for users to select a new results table; choose the measurement type into which those results will be uploaded; reapply default mapping; validate the data; then upload the new data for publication to the data lake, either automatically or with minimal input for polishing.

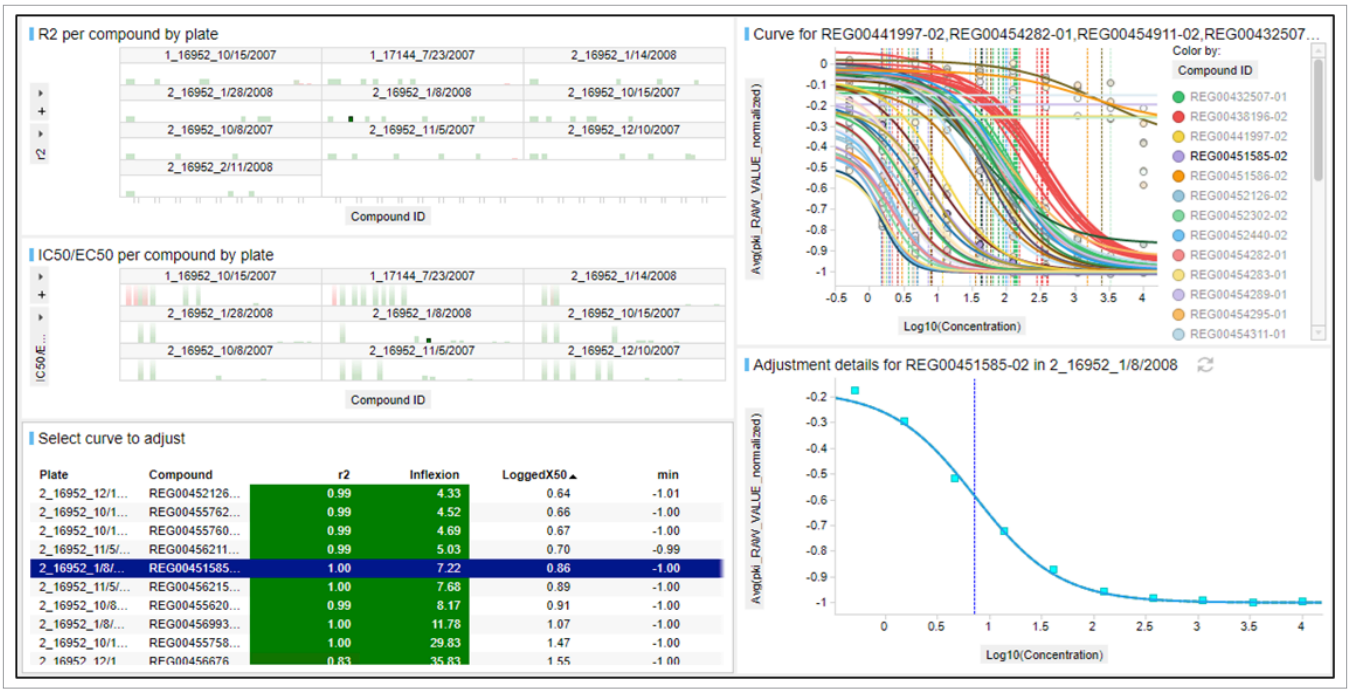

Figure 2. Signals Screening-Basic Screening curve fitting app provides plate by plate views of R^2, IC_{50}, and tabulated views of Log IC_{50}, min, max, and Hill coefficients alongside curve fits plotted onto the data. Visualizations are interactive such that selection of data point of interest highlights other visualizations. Data points can be excluded and the curve re-fit.

Signals Screening for HCS

Signals Screening-HCS processes HCS data similarly to HTS data. The ease with which algorithms can be utilized in Signals Screening-HCS is a key advantage. A majority of published HCS data exploits only one or two parameters in its analysis, meaning that this data is low content. Hence, researchers are not realizing the full capabilities of multiparametric datasets. By putting sophisticated multiparametric analysis tools into the hands of users, Signals Screening-HCS allows them to derive value from all their data.

Images from HCS instruments must first be subjected to an image segmentation process that translates them into numbers for analysis. Segmentation algorithms recognize morphological characteristics of cells and the location of subcellular features such as the nucleus, cytoplasm, and organelles. PerkinElmer's Columbus Image Data Storage and Analysis System, accessed via the Columbus Navigator app in Signals Screening, runs pre-designed segmentation routines and measures standard intensity and morphology parameters for translating image data into flat files. Columbus also contains sophisticated proprietary algorithms for quantitating texture or advanced morphology-like symmetry and compactness. The advantages of the Columbus system are transferable to the translation of image data from any manufacturer's HCS instruments.

Other common software for image segmentation/translation can also be used to migrate flat files into the Signals Screening environment via the Data Import app. The Plate Editor app is then used to assign metadata values to the plate, e.g., the location of positive and negative controls, compound concentrations, etc. Next, the user's desired data normalization method is selected within the Normalization app used for HTS data. As with HTS workflows, users requiring non-standard normalization methods can set these up in the Custom Calculation app. Once the data has been suitably normalized, any one of several available analytical regression methods in the Curve Fitting app can be applied to fit dose-response and other curves. After inspecting the curves to evaluate R^2 values, users can adjust asymptotes and remove outliers for a recalculation that may result in a more accurate ranking of phenotypic hits.

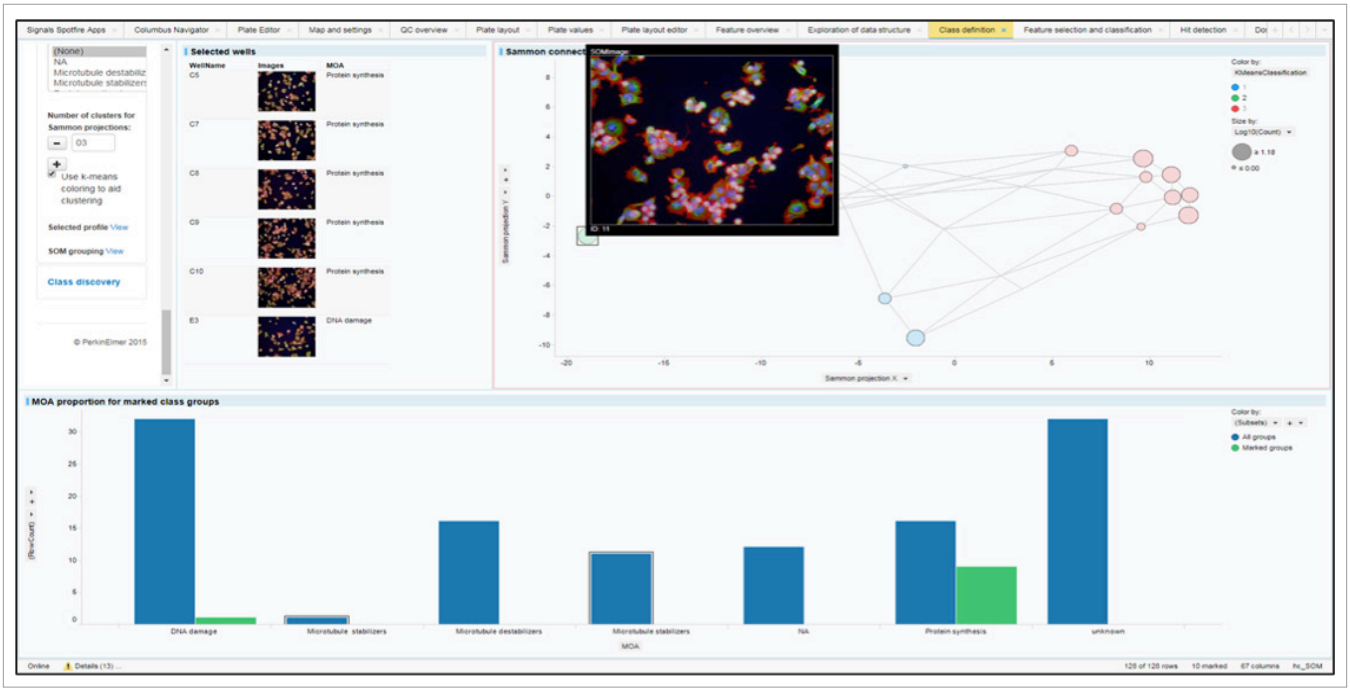

Figure 3. Creation of Self-Organizing Map (SOM) groupings in Signals Screening-HCS provides an unsupervised machine learning algorithm using data table features to group samples based on phenotypic profile. Similar phenotypes are clustered; relationships can be highlighted with a k-means function. An image rendering function allows users to check different phenotypes directly on the raw image within TIBCO Spotfire®.

HCS is the tip of the spear in biopharma for detecting phenotypic changes at the cellular level. Tracking and measurement of complex and subtle changes requires identifying which parameters best define a distinct cellular fingerprint. This type of reduction of dimensionality has traditionally occurred in one of two ways:

1. Users would rely on experience and intuition for deciding which features/components were most key
2. Or, the image files would be handed over to a bioinformatics team for data crunching.

However, the High Content Profiler and t-distributed stochastic neighbor embedding (t-SNE) apps in Signals Screening 2.0 give users the ability to make these determinations by the straightforward use of powerful multivariate statistical methods. This includes t-SNE, and unsupervised machine learning techniques such as Self Organizing Maps (SOM) and Principal Components Analysis (PCA). In this way, Signals Screening not only provides users more control over the data analysis process, but also saves time by eliminating the wait associated with handing off to the statistics department. Unsupervised machine learning can be used to create SOM groupings that define classes based on a certain phenotypic profile. These data analyses can also be captured as a protocol to automate work on subsequent datasets.

Signals Screening for SPR

Integration of Signals Screening-SPR within the TIBCO Spotfire® analytics platform offers tools for processing and analyzing SPR data from popular SPR instruments. Signals Screening-SPR can import Biacore T200 raw data files, Biacore 4000 data files, IBIS MX96 raw data files, and ForteBio Octet BLI data files. Other data formats will be added by PerkinElmer software developers based on customer demand.

SPR data analysis apps are located within their own area in the Signals Screening apps. After clicking Data Import to upload instrument files, users select Data Preprocessing to access a menu of standard techniques: zeroing, cropping, alignment, referencing, and blank subtraction.

Using the Kinetic Analysis app, kinetic rate constants can be solved for global analysis fitting by 1:1 interaction with mass transfer. Global or local R_{max} and refractive index (RI) can be selected for fitting. The resulting curve fits are plotted onto sensorgrams and parameters: k_a, k_d, K_D, k_t, R_{max}, RI, Chi^2 can be selected from an interactive data table. Plots of the residuals are presented to inspect the fit of the data. Steady state analysis is available for data that have reached equilibrium. Analyzed data can be exported to Excel, PowerPoint, PDF, etc., for presentations and reporting.

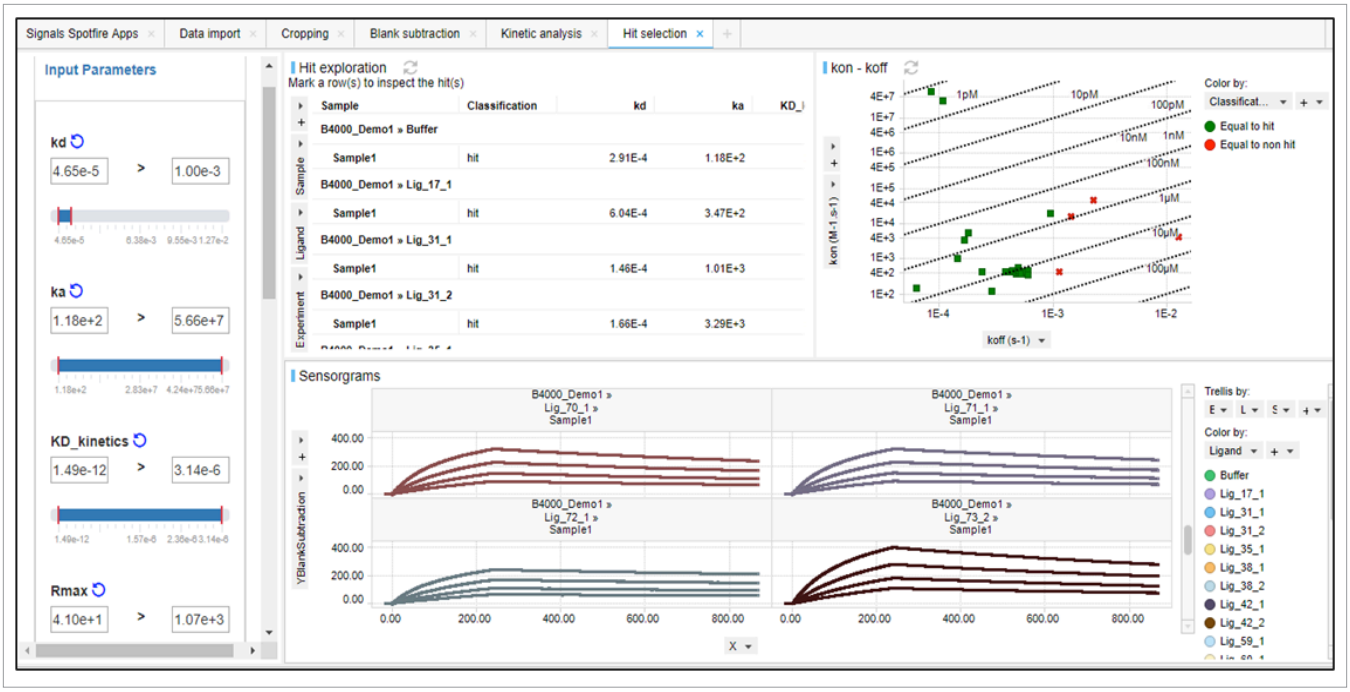

Figure 4. Signals Screening-SPR Hit Selection App contains an interactive isoaffinity plot to select hits of interest for sensorgram inspection and the ability to filter hits based on k_a, k_d, K_D or R_{max} values (left panel).

Signals Screening-SPR delivers expanded graphical capabilities and interactive visualizations capacities beyond most SPR instrument-specific software packages. For example, the Signals Screening-SPR Hit Selection app allows users to filter data by ka, kd, KD, and/or Rmax to select hits and inspect compounds that display desired binding characteristics. This filtered data is displayed in an isoaffinity plot alongside the corresponding sensorgrams. Conversely, data points of interest can be directly selected in the isoaffinity plot for inspection of the sensorgrams.

Use of a single program to analyze, visualize, graph, and report data from several popular SPR instrument platforms promotes standardization of SPR data analysis across the enterprise and eliminates cutting and pasting between programs, reducing human error and improving data integrity. These workflows can be saved as protocols for reuse in subsequent sessions, saving additional time, and increasing the consistency of SPR analysis results from individual scientists and within/across groups.

Summary

In the past drug discovery involved compromises that caused bottlenecks, inefficiencies, wasted time, and money. Among these, for many years, screening solutions were locked into a "closed box" approach to data processing. Scientists had to resort to bespoke tools, tightly coupled to screening instrumentation, and with limited flexibility.

PerkinElmer Signals Screening is an intuitive, configurable, and flexible screening workflow engine on top of the unparalleled data visualization and analysis capabilities of TIBCO Spotfire®. Signals Screening currently addresses three types of assays:

1. High Throughput Screening
2. High Content Screening
3. Surface Plasmon Resonance

Table 1. Current Capabilities of Signals Screening at a Glance.

	Inputs	Analysis Capabilities	Outputs
HTS	Generalized instrument file support of any text file and all standard plate readers	Normalization 3PL, 4PL, and linear curve fitting Curve fit statistics Fix asymptotes Remove outliers and re-fit Nonlinear dimensionality reduction: t-SNE	EC_{50}, IC_{50}, POC R^2, Hill coefficient, Stdev, Publish to Signals Lead Discovery Export to .ppt, .pdf, .csv, etc.
HCS	Instrument-agnostic, export results from common HCS instruments - Opera Phenix, Operetta CLS, CellnSight, etc. Direct import of Columbus image analysis results from within TIBCO Spotfire® Re-usable, easy to define import templates to bring in data from other image analysis tools like CellProfiler™, ImageJ/Fiji, Definiens and all other text-based results	Automated workflow Variety of QC options Exclude data points and re-analyze Normalization Curve fitting Principal Component Analysis (PCA) Unsupervised machine learning Single cell analysis Feature selection Classification	Hit selection Subpopulations based on phenotypic similarity Positivity score EC_{50}, IC_{50}, R^2 Box-plot and density plot for single cell statistics Plate heatmaps Export to .ppt, .pdf, .csv, etc.
SPR	Biacore T200 raw data Biacore 4000 partially ForteBio OctetRED IBIS MX96	Aligning Referencing Blanking 1:1 interaction with mass transfer global analysis fitting Steady State Analysis MW Normalization	k_a, k_d, K_D, R_{max}, K_t, x^2, stdev Sensorgrams with curve fit overlay; residuals Isoaffinity Plot Hit selection Export .ppt, .pdf, .csv, etc.

PerkinElmer, Inc.
940 Winter Street
Waltham, MA 02451 USA
P: (800) 762-4000 or
(+1) 203-925-4602
www.perkinelmer.com

PerkinElmer®

For a complete listing of our global offices, visit www.perkinelmer.com/ContactUs

Copyright ©2018, PerkinElmer, Inc. All rights reserved. PerkinElmer® is a registered trademark of PerkinElmer, Inc. All other trademarks are the property of their respective owners.

014282_01 PKI