The Chemical Biology Core Facility is a service unit at EMBL providing screening services to research groups at DKFZ, University of Heidelberg and EMBL and helps them to identify small molecules in a library of approximately 80,000 compounds which can be used as biotools to address important biological questions on novel targets.

Spotfire®

"For visualization and filtering of large and complex data, Spotfire is the only data analysis tool I can recommend because it actually does much more."

Jan Selig, Senior Technical Officer, European Molecular Biology Laboratory

TIBCO Spotfire® Analytics Software Helps to Identify and Develop Chemical Compounds In Academic Screening Facility

EMBL's Chemical Biology Core Facility – a joint venture between the European Molecular Biology Laboratory , the German Cancer Research Center(DKFZ) and Heidelberg University – successfully deploys Spotfire software's interactive and robust analytics capabilities

DEPLOYMENT SUMMARY

EMBL's Chemical Biology Core Facility uses TIBCO Spotfire analytics to analyze and decide on all crucial points in its high-throughput screening processes. Spotfire software's automated quality control helps to increase confidence in "hit" compounds, resulting in greater cost and time effectiveness.

The Chemical Biology Core Facility processes screening requests for small chemical molecules, utilizing a carefully chosen screening library of nearly 80,000 commercially available compounds. The aim is to address important biological questions by identifying and developing biotool compounds against novel targets e.g. to elucidate pathways or mode of action of binding partners.

When a screening request comes into the facility it is first evaluated for suitability and amenability by the facility's steering committee. Upon positive review a project manager chooses the appropriate screening platform and assay technology. Together with the requester the project manager develops an assay which is then adopted, optimized and automated to perform the high-throughput screening campaign. If an initial prescreen of over approximately 4,000 compounds shows no results the assay might have to be optimized or even redeveloped. In case of positive results the screening of the complete collection can be skipped and hit characterization can start immediately. In most cases though, the full screen is conducted followed by hit confirmation and characterization. To confirm hits, positives are selected and retested. For compound characterization (potency or toxicity) confirmed hits are tested in serial dilutions in dose response experiments to determine the half maximal effective concentration (EC50). Finally, analogues of the active structures can be identified to further explore structure activity relationships.

The project manager reports on the project progress and together with the customer, decides on the next steps.

TIBCO®
The Power of Now®

Challenge

While the introduction of high-throughput techniques and lab automation helped to save a lot of time and manual effort to screen compounds, organizations lacked an easy to handle tool to perform accurate analyses that didn't require a highly skilled analyst. Moreover to cope with the tremendous amount of data, the scientists used to evaluate results on a per batch basis, skipping over information from the entire dataset thus missing opportunities to make better decisions.

A large volume of complex data has to be assessed throughout the workflow to achieve the goal of delivering high quality compounds. When the assay is developed and optimized it's not only the activity or inactivity of a given target that has to be considered, batch size, reagent preparation, temperature and signal stability need to be evaluated for a robust assay. Moreover, the quality of the screening campaign needs to be supervised to eradicate systematic errors, and hits have to be carefully selected to minimize false positives and false negatives in order to concentrate on the most promising candidates and make the most out of the available resources.

Many data evaluation applications have difficulty dealing with the sheer volume of data, let alone handling the data's multivariate and multiparametric nature.

Systematic Errors per Batch uncorrected top and corrected below

Solution

The TIBCO Spotfire enterprise analytics platform was selected by the facility to analyze and support the presentation of large data sets and allow scientists to accelerate the screening process with powerful and interactive visualizations. EMBL Senior Technical Officer Jan Selig was familiar with Spotfire software's capabilities since he implemented it to perform microarray analysis prior to joining the Core Facility in 2004. Selig is now responsible for lab automation, data and knowledge management. "We choose Spotfire because it was the only tool capable of handling millions of data points with high performance and its simplicity to connect to and query data from disparate sources

including our oracle database out of the box," says Selig. "Spotfire gives us the opportunity to look at data from various perspectives at all the major steps within the project workflow. Especially when we develop the assay and after the screening runs we use Spotfire's visualization and filtering capabilities starting with a bird's eye view over the complete range of results and then drilling all the way down to a single data point and back."

Spotfire analytics is used by the Core Facility project managers to make critical decisions at key points in the HTS workflow. The Spotfire server – tightly integrated with the Core Facility's IBDS ActivityBase® data depository of chemical structures, plate content and experiment information – collects all of the data needed at each stage of the workflow. Very little customization was required because Spotfire analytics is readily adaptable for all types of analysis. "Spotfire's functionality gives me a lot of built-in options I didn't have before," notes Selig. "So I was able to do trellis plots, overlays, average calculations per position, data normalization and systematic error corrections, without having to do any coding."

In the development stage, the analytics platform is used to evaluate robustness of an assay and to choose optimum conditions for the large scale experiment. E.g. cross titrations of two reagents are performed to find the settings where minor changes in the underlying conditions cause only small changes for the dynamic range of the assay.

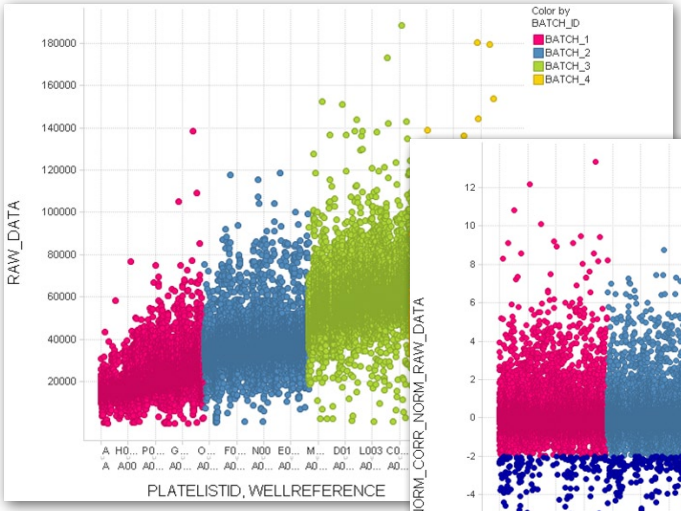
In the pilot screen phase of the primary screening stage, Spotfire analytics is used to analyze the initial 4,000 substances. The pilot screen usually is done in duplicate. The two sets are used to verify the chosen conditions. In addition, it helps us to practice for the full screen and get an idea about the hit rate that can be expected. Based on this initial screen, we report back and decide whether or not to proceed with a full screen of all 80,000 substances in the Core Facility's library.

It's this phase of full or lead screen of 80,000 substances that poses a real challenge for analysis tools. Spotfire analytics was one of the first to be able to process a massive and complex data set such as that used in HTS, and today remains one of the few tools that can do the positive selection job quickly, cost-effectively and with the quality control and accuracy needed for the successful screening of substances. The results of the full screen are directly loaded from the database into Spotfire analytics without preprocessing, and using pre-loaded templates and workflow, quality control measures can be performed.

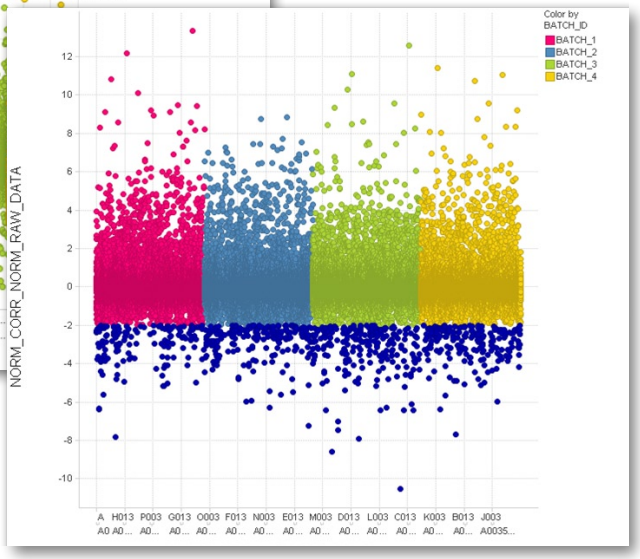
The project manager looks at the Z-scores, a measure for data quality that tells the project manager if there is an error in the assay assembly. Using standard scatter plots generated by Spotfire analytics the overall performance of an assay can be assessed over time. The software's visual interface makes this easy to spot. To normalize screening raw data and to compensate for systematic errors introduced by the automated liquid handling, reader devices and consumables a statistical calculation takes place inside Spotfire analytics. The results of the analysis and the changes applied to the data during correction can be visually followed with Spotfire analytics, making data processing transparent to the users.

"What impresses me most is that if you look at a cloud of data points in Spotfire analytics, you instantly recognize patterns you wouldn't see by just looking at raw data or plate based data evaluation," says Selig. "Since it's so easy to put data into Spotfire, we are now able to look at the results on various levels and from multiple angles. Once we see patterns within datasets we can immediately identify them and start to explore. We know for example that signals depend on incubation time and changes in temperature and we expect to see that the overall signal changes over the single plates of a batch," continues Selig. "What we don't expect is a recurrent pattern within the single plates. If we find these patterns within plates we know that they usually come from systematic errors introduced by the automated liquid handling devices, the readers or consumables." By looking at averages over the position of the plate, Core Facility project managers can detect systematic errors where they wouldn't have expected them. And, once detected, these systematic errors can be compensated for. "This wouldn't be as easy with any other tool," says Selig.

After the secondary screening Spotfire analytics prepares all data collected and enables the researcher/project manager to clearly identify which substances were found by the assay and whether they are inhibitors or activators of a target reaction. Spotfire analytics can also be used to export a report to PowerPoint, narrowing down the vast amount of data to something manageable that can be presented to the customer.



Raw Data Full Screen, 157 plates, 50000 compounds showing signal drifts over batch processing time



Normalized Raw Data Corrected and Renormalized, highlighted in blue selected positives for confirmation

Results

Spotfire analytics provides the Core Facility with a number of tangible benefits. "The biggest benefit for our group is that we can interactively analyze results in Spotfire and communicate these visualizations to our customers much easier than we could with any other tool," says Selig. "Moreover, with Spotfire's automated quality control we are confident that our results are reliable." Spotfire software's advanced statistical capabilities facilitate normalization and error correction and this functionality has enabled the Core Facility to improve the positives selection by increasing the confirmation rate. At the same time resources in subsequent follow up testing are saved because noise can be reduced as well as the wrong selection of compounds. Spotfire software's fast data retrieval and standard visualizations help to speed up the entire HTS process: project managers can evaluate the quality of a screening campaign minutes after the data upload, take appropriate actions when anything goes wrong and decide whether or not to proceed.

Spotfire analytics has proven to be highly cost-effective because the Core Facility could avoid costly follow-up by minimizing false positives. In addition, the facility can spare a dedicated data analyst because Spotfire analytics helps project managers to easily handle big data sets, and limits the data size to a manageable amount for final reports.

The Core Facility plans to use Spotfire software's robust interactive analysis capabilities for follow-up testing and cross comparisons between assays. This would help to look for compounds showing up in all assays that might be interfering with the technology, but not necessarily the desired target. "For visualization and filtering of large and complex data, Spotfire analytics is the only data analysis tool I can recommend, because it actually does much more" concludes Selig.

TIBCO®
The Power of Now®

TIBCO Software Inc. (NASDAQ: TIBX) is a leading independent business integration software company and a leading enabler of real-time business, helping companies become more cost-effective, more agile and more efficient. TIBCO has delivered the value of real-time business, what TIBCO calls The Power of Now®, to thousands of customers around the world and in a wide variety of industries.

For more information contact authorized Spotfire Partner:

SyncSite

5330 Stadium Trace Parkway
Suite 100
Birmingham, AL 35244

Tel: 1-205-443-1000
www.syncsite.net