Relating Structures to Targets Through the Analysis of Quality-controlled Large-scale Screening Data

Genedata
solutions in-silico

Michael Lindemann, Oliver Duerr, Tom Jung, Bernd Kappler, Rahel Luethy,
Swen Reimann, Christian Ribeaud, Bernd Rinn, Lee Sargeant, Tim Wormus, and Stephan Heyse of Genedata AG, Basel, Switzerland

High-throughput screening (HTS) operations systematically explore the space of drug-like molecules, measuring the activity of 100,000s of compounds on a variety of biological targets in standardized and well-controlled experiments, providing a wealth of data which has so far rarely been fully exploited.

At Genedata, we are developing sophisticated methods and software for the systematic, automated analysis of large-scale screening data. We summarize the workflow to identify relationships between structural clusters and their effects on target classes in the large set of bioactivity profiles from HTS and secondary screening campaigns.

As an important first step in our analysis process we perform rigorous quality assurance (Figure 1). Second, we process and standardize the resulting high-quality data both from single-dose and dose-response screens (Figure 2) and align them to meaningful bioactivity profiles. Third, we assess the global reproducibility and comparability of these data sets (Figure 3), before - fourth - target-unrelated compound activities are eliminated (Figure 4). Fifth, we relate the structural classes of compounds to their bioactivity and specificity by in silico compound profiling (Figure 5).

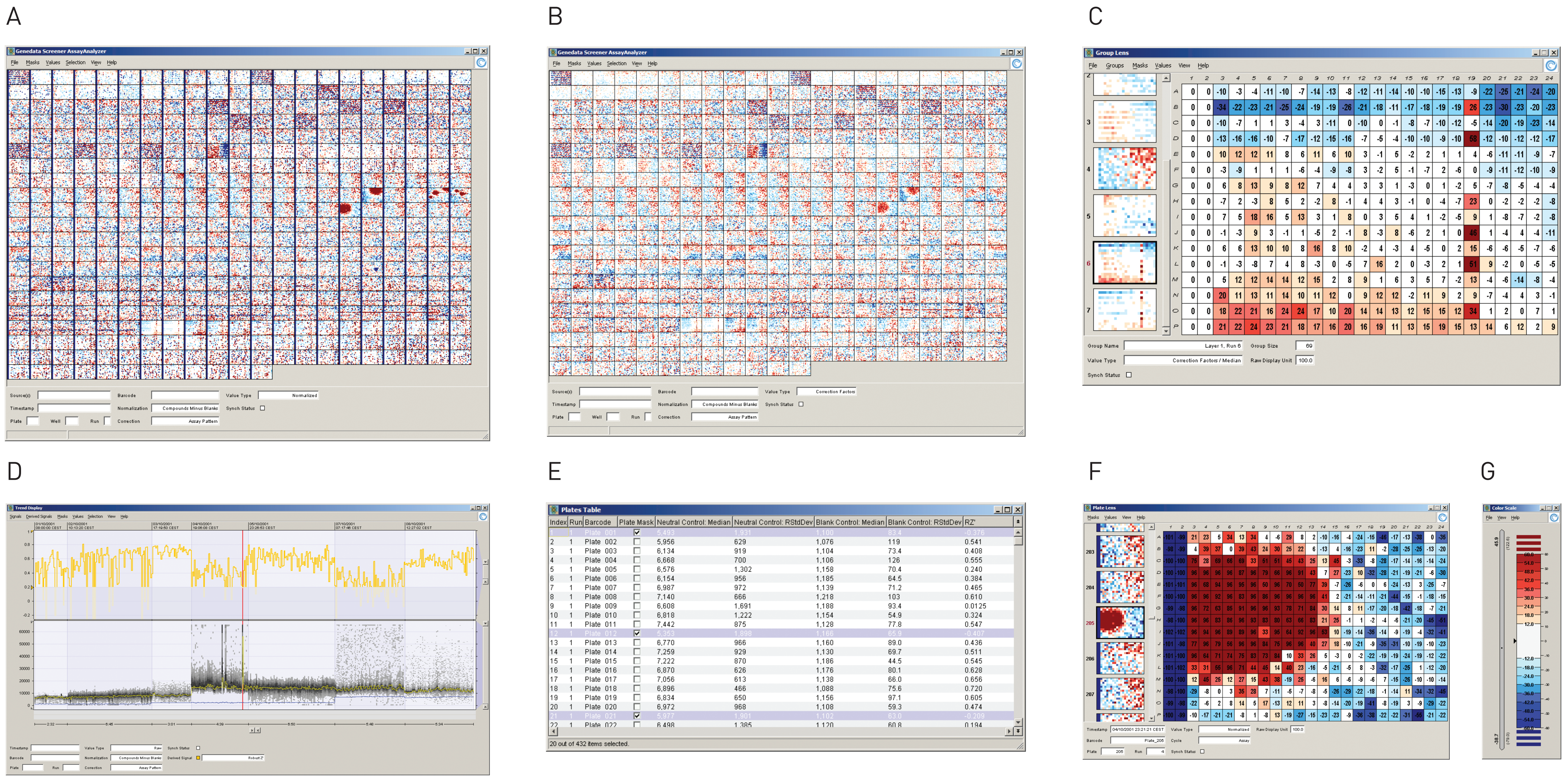
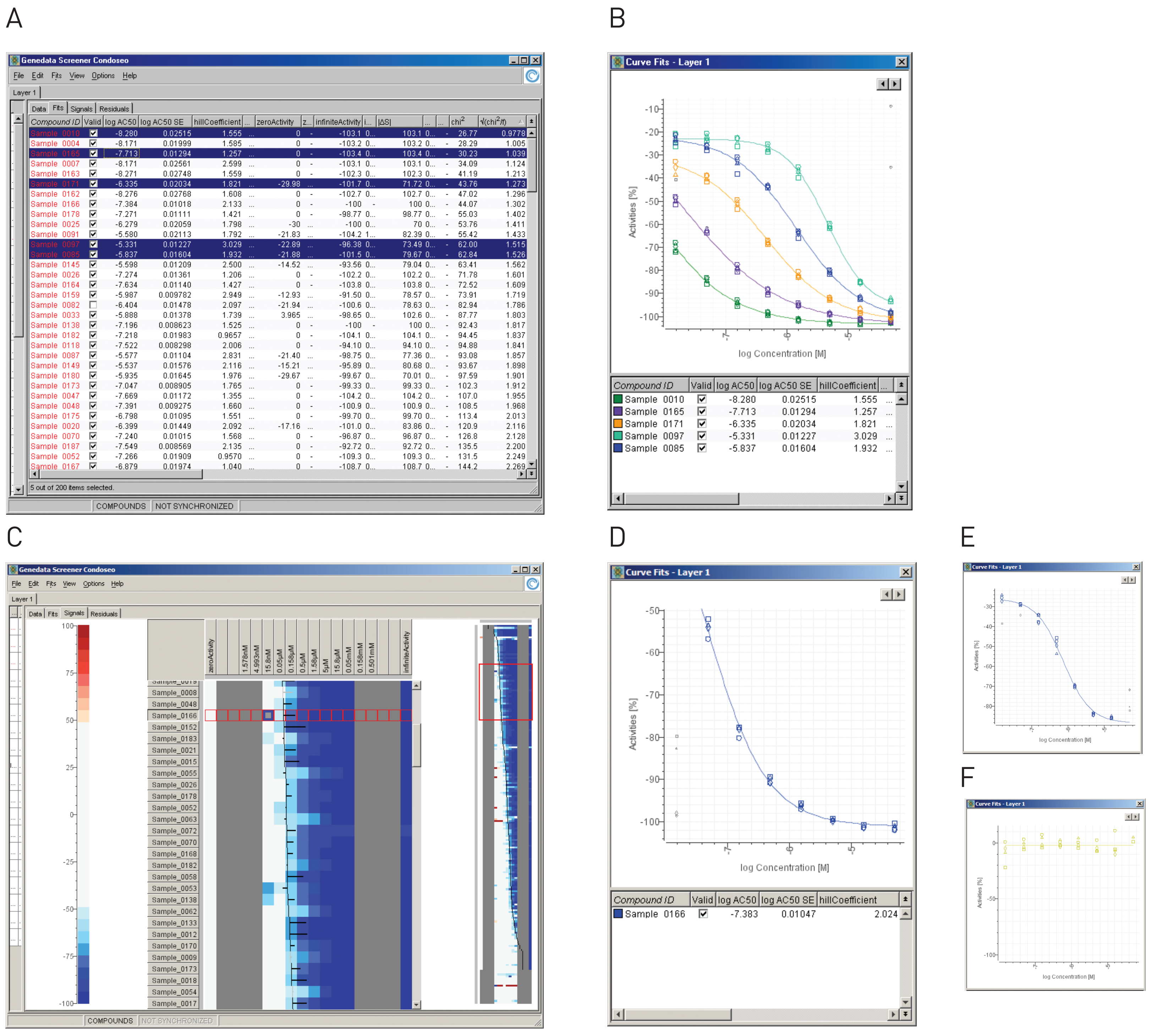


Figure 1

Screener AssayAnalyzer's powerful algorithms and efficient visualizations allow for a rapid identification of process problems
The overview in Panel A displays platewise normalized signals from an entire screening campaign. Plates (rectangles) are arranged in chronological order; wells are color-coded according to the signal scale given in Panel G. Panel D shows trends in the plate summary statistics (Z' factors, upper; raw signals, lower pane).These overviews allow detection of failed plates very efficiently (Panel E, plates with low Z' factor, Panel F, contaminated plate with extreme values). The pattern view (Panel B) reveals systematic process problems by highlighting the automatically detected, repetitive patterns in this assay. More subtle errors are detected using the group statistics. Panel C shows the well medians from all 69 plates in run 6 indicating systematic deviations in column 19 that persist in run 7 (thumbnail).



Compound ID	Valid	log AC50	log AC50 SE	hillCoefficient	...
Sample 0010	☑	-8.280	0.02515	1.555	...
Sample 0165	☑	-7.713	0.01294	1.257	...
Sample 0171	☑	-6.335	0.02034	1.821	...
Sample 0097	☑	-5.331	0.01227	3.029	...
Sample 0085	☑	-5.837	0.01604	1.932	...

Compound ID	Valid	log AC50	log AC50 SE	hillCoefficient
Sample 0166	☑	-7.383	0.01047	2.024

Figure 2

Screener Condoseo's smart curve-fitting algorithms and compact visualizations enable efficient, high-quality processing and analysis of dose-response-curve experiments.
Panel A shows the tabular overview of fit results after rapid fitting of all compound measurements and sorting by fit quality. Individual high-quality fits are displayed in the dose-response curve window (Panel B). Condoseo's sorting and selection capabilities, combined with concise displays such as the signals and fit parameter overview (Panel C) allow for an efficient assessment of hundreds and even thousands of dose-response curves. It's smart fit algorithms automatically identify and exclude outliers as shown for the measurements at the lowest concentration in Panel D (grey markers), and handle typical "fit problems" such as bell-shaped curves (Panel E) or inactive compounds (Panel F).

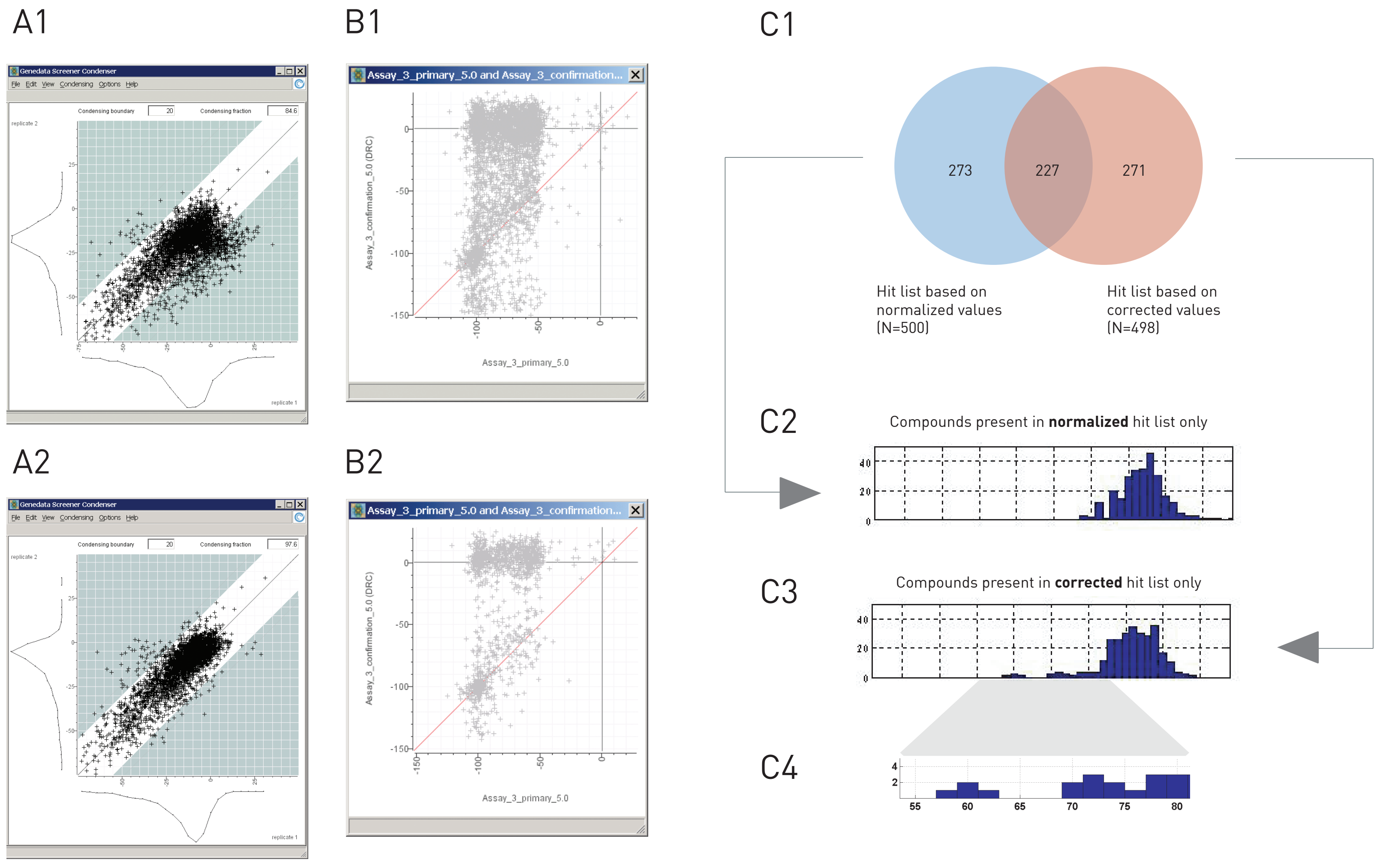


Figure 3

Rigorous quality control improves reproducibility and significance of assay data .
The scatter plots in Panels A1 and A2 compare replicate measurements for a set of compounds in one HTS assay. Panel A1 shows the result for normalized activity data, Panel A2 the result for quality-controlled and corrected data obtained by applying Genedata's proprietary error detection and correction algorithms. Comparing A1 and A2, the gain in reproducibility - and thereby significance - is evident.
Panel B compares the outcomes of a single-dose HTS screen and a dose-response-curve validation screen on the same compound set. Both axes represent activity values at the primary screening concentration; for the validation screen the activity was back-calculated from the dose-response-curves. In Panel B1 all data is taken into account, yielding a diffuse point cloud obscuring any relationship between the two assays. In B2 only the data with low standard error of the IC50 estimate is used. This high-quality data set shows a match between primary and validation screens for the true actives and separates these from the false-positives of the primary hit selection.
Panel C summarizes the result from a case study, comparing hit selection on normalized screening data vs. data subjected to an additional quality assurance step. Panel C1 shows the overlap of the 500 most active compounds selected in each scenario. Panel C2 depicts the activity distribution in the validation screen for compounds only present in the hit list based on normalized data, Panel C3 the distribution based on corrected data. The enlarged view in Panel C4 shows how the correction of systematic deviations can rescue false-negatives.