## Relating Structures to Targets Genedata solutions in-silico Through the Analysis of Qualitycontrolled Large-scale Screening Data

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.

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 - targetunrelated 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).

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.



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).

Trend Display												
Signals Derived Signals Masks Values Selection View Help												
01/10/2001 02/10/2001 08:00:00 CEST 10:13:20 CEST	03/10/2001 17:19:50 CEST	04/10/2001 19:06:08 CEST	05/10/2001 23:26:53 CEST	07/10/2001 07:17:46 CEST	08/10/2001 12:27:02 CEST							
				0 0 -	when he had he h	*						
				MAR MAR NO		-						

ndex	Run	Barcod	e	Plate Mask	Neutral Control: Media	n Neutral Control: RStdDev	Blank Control: Me	dian Blank Control: RStdDev	RZ' 4
	1	Plate C	01		5,493	1,931	1,100	83.4	-0.376
2	1	Plate 0	02		5,956	629	1,076	119	0.541
3	1	Plate 0	03		6,134	919	1,104	73.4	0.408
1	1	Plate 0	04		6,668	700	1,106	126	0.555
5	1	Plate 0	05		6,576	1,302	1,158	70.4	0.240
6	1	Plate 0	006		6,154	956	1,185	64.5	0.384
7	1	Plate 0	07		6,987	972	1,139	71.2	0.465
3	1	Plate 0	80(		7,140	666	1,218	103	0.610
3	1	Plate 0	09		6,608	1,691	1,188	93.4	0.0125
0	1	Plate 0	)10		6,818	1,222	1,154	54.9	0.324
11	1	Plate 0	)11		7,442	875	1,128	77.8	0.547
13	1	Plate 0	)13		6,770	966	1,160	89.0	0.436
4	1	Plate 0	)14		7,259	929	1,130	69.7	0.511
15	1	Plate 0	)15		7,222	870	1,186	44.5	0.545
16	1	Plate 0	)16		6,870	626	1,176	80.1	0.628
17	1	Plate 0	)17		7,056	613	1,138	66.0	0.656
18	1	Plate 0	)18		6,896	466	1,088	75.6	0.720
9	1	Plate 0	)19		6,834	650	1,156	97.1	0.605
20	1	Plate 0	20		6,972	968	1,108	59.3	0.474
12	1	Plate (	122		ñ 498	1.385	1 1 2 0	60 B	0 1 9 4





🥘 G	Cenedata Screener Condoseo															
File	<u>E</u> dit F <u>i</u> ts	: ⊻iew	<u>O</u> ptions	<u>H</u> elp												O
Layer 1																
	Data Eit	s Ì sians	ale Ì Resid	( alouk												
<b>1</b>	Data	- Joigne www.h	ais į Resiu	uais [		hill O a set si sut	_			ميا يا.	01	ГГ	T.	2	200	
	Compou			ACSU	IUG ACSU SE	millooenicient		zeroActivity Z	ImmileActivit		ioj 402.4		[0	00.77	A(CHI A)	0770
	Sample	0010		-8.280	0.02515	1.555		U ·	-103.	1 0	103.1	0	-	20.77	U	9778
Ľ.	Sample	0004		-8.171	0.01999	1.585		U ·	· -103.	2 0	103.2	0	-	28.29		1.005
	Sample	0007		0.171	0.01294	1.207		0 •	-103.	4 0	103.4	0	-	30.23		1.1.24
Ë I	Comple	0007	<b>₩</b>	-0.171	0.02561	2.099		0 -	-103.	2.0	103.1	0	-	34.09		1.124
	Sample	0103		-6.271	0.02748	1.008		- 20.09	-102.	30 70	71.72	0	-	41.18		1.213
H I	Sample	0162		-8.276	0.02034	1.608		-23.30	-101	7 0	102.7	0		43.70		1 296
ŀ	Sample	0166		-7.384	0.02100	2 1 3 3		0 -	-10	0 -	100	0	-	44.07		1.302
	Sample	0178		-7 271	0.01111	1 421		0 -	-98.7	- 7 0	98.77	0	-	55.03		1 402
·	Sample	0025	V	-6.279	0.02059	1.798		-30	-10	0.0	70	0	-	53.76		1.411
	Sample	0091	<b>V</b>	-5.580	0.02113	1.792		-21.83	-104.	2 1	82.39	0		55.42		1.433
	Sample	0097		-5.331	0.01227	3.029		-22.89	96.3	8 0	73.49	0	-	62.00		1.515
	Sample			-5.837	0.01604	1.932		-21.88	101.	5 0	79.67	0		62.84		1.526
-	Sample	0145	<b>v</b>	-5.598	0.01209	2.500		-14.52	93.5	6 0	79.04	0	-	63.41		1.562
	Sample	0026	~	-7.274	0.01361	1.206		0 -	-102.	2 0	102.2	0	-	71.78		1.601
- 1	Sample	0164	~	-7.634	0.01140	1.427		0 -	-103.	8 0	103.8	0	-	72.52		1.609
	Sample	0159	<ul> <li>Image: A start of the start of</li></ul>	-5.987	0.009782	2.949		-12.93	91.5	0 0	78.57	0	-	73.91		1.719
	Sample	0082		-6.404	0.01478	2.097		-21.94	100.	6 0	78.63	0	-	82.94		1.786
ŀ	Sample	0033	✓	-5.888	0.01378	1.739		3.965	98.6	5 0	102.6	0	-	87.77		1.803
	Sample	0138	<b>~</b>	-7.196	0.008623	1.525		0 -	-10	0 -	100	0	-	92.43		1.817
·	Sample	0182	✓	-7.218	0.01983	0.9657		0 -	· -104.	1 0	104.1	0	-	94.45		1.837
	Sample	0118	$\checkmark$	-7.522	0.008298	2.006		0 -	-94.1	0 0	94.10	0	-	94.88		1.841
Ŀ	Sample	0087	✓	-5.577	0.01104	2.831		-21.40	98.7	5 0	77.36	0	-	93.08		1.857
	Sample	0149	✓	-5.537	0.01576	2.116		-15.21	95.8	9 0	80.68	0	-	93.67		1.898
- 1	Sample	0180	~	-5.935	0.01645	1.976		-29.67	99.6	7 0	70.01	0	-	97.59		1.901
	Sample	0173	~	-7.047	0.008905	1.765		0 -	-99.3	3 0	99.33	0	-	102.3		1.912
F I	Sample	0047	~	-7.669	0.01172	1.355		0 -	-104.	2 0	104.2	0	-	107.0		1.955
	Sample	0048		-7.391	0.009275	1.660		0 -	· -100.	9 0	100.9	0	-	108.5		1.968
F I	Sample	0175		-6.798	0.01095	1.551		- 0	-99.7	0 0	99.70	0	-	113.4		2.013
·	Sample	0020		-6.399	0.01449	2.092		-17.16	101.	0 0	83.86	U	-	120.9		2.116
	Sample	0070		-7.240	0.01015	1.568		U -	96.8	7 0	96.87	0		126.8		2.128
Ŀ	Sample	0187		-7.549	0.008569	2.135		0 -	-92.7	2 0	92.72	U	-	135.5		2.200
	Sample	0052		-1.200	0.01909	0.9570		0 -	· -109.	30	109.3	0	-	131.5		2.249
Ë	elame	0167		-0.879	0.01974	1.040		U -	-108.	/ 0	108.7	0	-	144.2		2.209
	E out of 2	D How o	oplastad				_									
Ľ	13 Out 01 2	50 items	selected.													
													_			







**B**2

















В



## 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 doseresponse 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

A1

## 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.

Assay\_3\_primary\_5.0

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-responsecurves. 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.