

HCS Analyzer tutorial v1.0.4.1

This tutorial illustrates HCS analyzer functionalities, showing how the user can handle the analysis of an entire High Content Screening exported database.

For this tutorial, we used an artificially generated siRNA screening database composed of 25473 wells distributed over 68 plates. The plate format is a regular 384 wells plate, with 24 columns and 16 rows.

Content

- 1. Import Screening
- 2. Define Controls (if present)
- 3. Quality control
- 4. Descriptor Selection
- 5. Phenotypic clustering
- 6. Phenotypic classification
- 7. siRNA screening
- 8. Development
- 9. Plugin Development



Import Screening

Even if HCS analyzer is able to import MTR files, it has been principally designed to load CSV file. In such file 2 columns are mandatory: one column containing the plate name and another one with the well absolute position of the well. Each row represents a well where any descriptors can be added in a **double** format. The user has the possibility to add 3 others specific data:

- A name (compounds or siRNA name by example) as a string
- An information as a string
- A Locus ID as an integer that will allow the software to make the link to KEGG database.

Here is a typical example of such file generated by MS Excel and save in CSV format.

<u>Note</u>: The names as well as the position of the column do not have to follow any rule as the software loader is flexible enough to deal with this kind of variability.

Plate 384	Well 384w	Cat	Catalog Numt Gene Symbol	Description	Locus ID	Fluo1	Size	Volume	Ratio1	Fluo2	Texture	Rand1	Rand2	Rand3	Rand4
SRP000459	D24	Druggable	1	1 adrenocortical dysplasia homolog	65057	0.822859	0.982368	59.26912	144.2518	90.04089	104.5234	48.66298	35.7505	36.28192	14.83044
SRP000479	P20	Druggable	2	2 deleted in esophageal cancer 1	115123	0.72335	0.987039	183.9027	234.0327	176.5919	231.5573	0.90994	7.755161	22.86228	23.36147
SRP000502	G11	Druggable	3	3 hypothetical protein CG003	57574	0.909086	0.724062	137.0043	139.0109	74.81597	195.0901	34.82006	5.351135	40.45101	0.397935
SRP000469	P09	Druggable	4	4 hypothetical protein 15E1.2	54708	0.75825	0.943229	58.89768	196.4233	54.3536	118.3427	35.54298	47.71962	46.30442	9.891081
SRP000463	F05	Druggable	5	5 82-kD FMRP Interacting Protein	10299	0.913974	0.936443	144.9422	481.5817	40.2687	156.4878	33.01084	18.35786	16.16318	32.8679
SRP000474	K16	Druggable	6	6 hypothetical protein 384D8_6	64844	0.951024	0.940594	22.85816	499.7528	7.579783	113.2114	3.485309	5.603355	46.58561	42.21598
SRP000480	D10	Druggable	7	7 3' exoribonuclease	92979	0.983895	0.993165	23.72045	125.6528	19.88528	38.27095	17.06228	42.23741	13.09605	14.63739
SRP000466	E13	Druggable	8	8 gamma tubulin ring complex pro	23591	0.781394	0.883409	288.7793	1185.202	112.0497	310.5889	35.11503	24.39241	49.79913	9.485995
SRP000439	D16	Druggable	9	9 putative binding protein 7a5	1731	0.941405	0.86835	180.8949	209.4964	103.6978	140.4776	10.84225	7.070599	28.18577	36.48984
SRP000469	M24	Druggable	10	10 hypothetical protein FLJ13511	54464	0.912182	0.935283	146.6728	523.9838	132.9783	224.7909	6.941006	2.267799	18.93494	38.59139
SRP000439	D18	Druggable	11	11 8D6 antigen	55964	0.916994	0.882256	155.6142	187.1925	54.14138	91.08958	38.03543	4.182601	11.05194	11.11748
SRP000439	D20	Druggable	12	12 alpha-1-B glycoprotein	23157	0.917484	0.791711	154.0979	120.4616	122.1243	100.1224	13.09587	14.02423	6.747342	1.986987
SRP000465	108	Druggable	13	13 ataxin 2-binding protein 1	23176	0.937335	0.848266	113.7	171.2941	25.2004	33.61339	3.128092	1.025906	6.887327	31.69431
SRP000484	121	Druggable	14	14 ataxin 2-like	151011	0.861691	0.974512	61.80003	157.4689	96.43404	89.46256	47.26814	33.45952	5.308625	29.8281
SRP000438	A05	Druggable	15	15 alpha-2-macroglobulin	55752	0.347707	0.899089	67.99509	159.2583	90.66948	154.9841	25.47527	14.39126	1.822695	33.67102
SRP000483	M07	Druggable	16	16 alpha 1,4-galactosyltransferase	143501	0.727905	0.732156	173.724	630.3225	146.8105	148.0179	44.21375	14.64555	22.50402	47.52822

<u>Note:</u> HCS analyzer can deal with two different well positioning modes that can be selected in the options window.

Options										
Correlation matri	x Misc	1								
Import / Export	Display	Normalization	Clustering	Classification						
-Well position n	node									
	A01		1.1							
•	Single Mode									
Import Names										
		not named, the	n inactive							
		Ok								

1- Single Mode: The row position is given by an alphabetical character. Row and column positions are merged together in a single column.

2- Double Mode: One column has to be created for the row position and one for the column.

Regarding the selected mode, the loading window options will be adapted.



First, click on "Import Screen" from the "File" Menu, and locate the file to be loaded.



Plate Co	Dimensions Ilumns 24	A. V						
Ro	ows 16	×						
	Data Name	Selection	Туре		Readout 0	Readout 1	Readout 2	Readout 3
	Plate384		Plate name	-	SRP000459	SRP000479	SRP000502	SRP000469
	Well 384w		Well position	-	D24	P20	G11	P09
Ø	Cat		Info	-	Druggable	Druggable	Druggable	Druggable
	Catalog Number		Descriptor	-	1	2	3	4
	Gene Symbol		Descriptor	-	1	2	3	4
	Description		Name	•	adrenocortical dy	deleted in esoph	hypothetical prot	hypothetical prot
	Locus ID		Locus ID	-	65057	115123	57574	54708
	Fluo1		Descriptor	-	0.8228594	0.7233504	0.9090863	0.7582496
	Size		Descriptor	-	0.9823682	0.9870385	0.7240621	0.9432293
	Volume		Descriptor	-	59.26911984	183.9027118	137.0042838	58.89768267
	Ratio 1	V	Descriptor	-	144.2518279	234.0326818	139.0108669	196.4233229
	Fluo2		Descriptor	-	90.0408859	176.5919403	74.81596763	54.35359857
	Texture	V	Descriptor	-	104.5234322	231.5573299	195.0901458	118.3426772
	Rand1		Select All		297765	0.909940097	34.82005831	35.54297666
	Rand2		Unselect /	All	5050135	7.755161101	5.351135178	47.7196201
	Rand3		Descriptor	-	36.28191786	22.86227618	40.45101283	46.30441893
	Rand4		Descriptor	-	14.83044153	23.36147298	0.397935311	9.891080863

The Import window should popup such as below. The software reads 4 rows to help the user to define its choice.

Define the plate dimensions, select all the requested descriptors as well as the other useful information such as Locus ID, Info or Name (if needed) click "Ok". A message box should appear giving you the information about the number of loaded well.



<u>Note:</u> In order to avoid any trouble during the process, the software does not accept any undefined well. It means that if any readout is missing, the concerned well will be automatically rejected.

Define Controls (if present)

Positive and/or negative controls have been performed during the screening process; the user can define them manually. First choose the desired class using the dedicated control, then using the mouse select the corresponding wells.



Display class mode

Note: It is easier to see the selection by switching to the "Display Class Mode".

<u>Note:</u> By default "Apply to all plates" is checked. Thus the current selection will be performed on every active plate.

Quality control

Many different way control the quality have been implemented in HCS analyzer. The Z-factor is a commonly used one that the user can test either on the current plate by clicking on "Plate->Quality Controls->Z-Factor" or "Screen->Quality Controls->Z-Factor".







In the first case, the Z-factor is computed on the current plate for each active descriptor and the results are sorted from the highest to the lowest. If performed on the complete screen, only the active descriptor is taken into account, and the associated Z-factor is represented for all the active plates allowing a rapid estimation of the low quality plates. Based on that, the user can reject the plate with a Z-Factor higher than a certain threshold by clicking on the "Reject Pates" button.

🚣 HCS analyzer v1.0.3		
File Edit SRP000459	 Plate Screen View Plug-ins Help 	
Current Plate Demensionality Reduction Systematic error ide	Prince Suber View Playins Teep Playing Teep Playins Teep Playing Teep P	Class selection Postive (0)
Systematic error identification	Plate by Plate Correction Reject Plates	S

The active plate list has then been modified, and a new Z-factor visualization gives:





The user can always display and modified the plate list by clicking on "View -> Plates Manager". The left column represent the complete set of loaded plates and the right column the set of active plates that will be processed.



In certain situation, assessing the normality of a descritpor readout on a control set of wells can also provide a precious information about the data quality. To check this, the user will select a descriptor, a class and then click either on "Plate -> Visualization -> Histogram"





<u>Note:</u> The histogram display and computation is highly modifiable through the "Edit->Options" window.



or "Plate -> Quality Controls -> Normal Probability Plots".

<u>Note:</u> Those displays can also be generated for complete screen by going to "Screen" menu.

If the user is interested by assessing the quality of a particular set of well among the complete screen, he can click on "Screen->Quality Controls->Descriptor Evolution" and check the evolution of the average values and standard deviations of the selected class.





Descriptor Selection

The presented application is clearly targeting screening multivariate data. In such context one important step in the data analysis is identify redundant source of information. Typically the identification of two correlated descriptors has to be identified either to perform better in the classification or clustering step, but can also be of great interest in understanding some biological mechanism of action.

If the user is interested by the quality of the descriptors in term of information, he can click on "Plate->Quality Controls->Correlation Matrix" (or its equivalent "Screen->Quality Controls->Correlation Matrix" to display the correlation matrix of the active descriptors). The process will display first the correlation matrix (by default the absolute value of the Pearson coefficient),



Note: For this process all the activated wells are taken into account (no specific class).

then the ranking of the coefficient to easily identify correlated coefficients.



<u>Note:</u> The user can select the Spearman coefficient and remove the ranking display by changing the options in the "Edit->Options" menu.



HCS Analyzer provides different approaches for automated feature selection. Those methodologies are available by clicking on the "Dimensionality Reduction" tab.

🚣 HCS analyzer v1.0.3			- • • ×
File Edit SRP000459	 Plate Screet 	en View Plug-ins Help	
Current Plate Dimensionality Redu	ction Systematic Error Identification & Corre	ction Amailzation Classification & Oustering Report Export	Class selection Positive (0)
New Dimension	2		Current Descriptor
Unsupervised	Supervised		Descriptor List
PCA v	InfoGain v Neutral Class 2 v		V Fluo 1 V Size V Volume Ratio 1 V Fluo 2 V Texture
Principal Component Analysis. For more information, go to: http://en.wikipedia.org/wiki/Principal_component_ analysis	InfoGain. For more information, go to: <u>http://en.wikipedia.org/wiki/Information_gain_in_d_</u> <u>ecision_trees</u>		 ✓ Rand1 ✓ Rand2 ✓ Rand3 ✓ Rand4
		HS	
Reduce Dir	nensionality	anallyzər	

At this stage, one can choose between two different kinds of approaches: Unsupervised or Supervised. Compare to Supervised, Unsupervised dimensionality reduction methods do not required any predefined classes.

Unsupervised: In this version of the software, two unsupervised methods are available: PCA or Greedy Stepwise.



<u>Note:</u> whatever the feature selection algorithm is selected, a warning message box will appear prior to the process, reminding the user that the active descriptor list will be modified.





If the PCA algorithm is selected, all the active wells will be used to identified the most important descriptors (i.e. the features maximizing the deviance of the data), reducing the dimension to the selected value.

Descriptor List		Descriptor List
V Huo1 V Size V Volume Ratio 1 Fluo2 Texture Rand1 Rand2 Rand3 Rand4 Rand4	New Dimension 2	Fuo1 Size Volume Ratio1 Fluo2 Texture Rand1 Rand2 Rand3 Rand4

Unlike the PCA, the greedy stepwise does not reduce the number of feature to a user selected value. The final number of features is automatically chosen by the software.

Supervised: As said above, supervised methods require a pre-defined set of classes. Let's select two controls as class 0 and 1 and keep the rest of the plate as class 2.

H.	e HCs analyzer v1.0.3																									
Γ	File	Edit	SRP000	0480									• P	late :	Screen	View	Plu	ig-ins	Help							
		Curr	rent Plat	·	➡∟	Dimensio	onality R	eduction	י 🔽	<u>.</u> :	Systema	tic Error	Identific	ation & C	Correctio	n [<u> </u>	Normaliz	ation		Class	ification	& Cluste	ering		Report Export
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	0.997 Current Descriptor
	A				L																					Global Fluo1 -
	C		_		-													-								Descriptor List
	D		-		-						-		-				-	⊢	-	-			-			Global only selected
	Е																									Volume Ratio 1
	F																									Fluo
	G																									Apply to all plates
	H																									I III Rand2 IIII Rand3
	1																									Rand4
	ĸ		-	-	-						-		-				-		-	-			-	-	-	Display class
	L				-														-				-			
	м																									
	N																									0.061
	0																									
	P																									

The algorithms will then choose the features that separate the more efficiently the classes, but will not take into account the well corresponding to the neutral class. In such way, the user can focus on identifying important features required for maximizing the difference between controls only.

		Supervised
	InfoGain	•
Neut	ral Class	2 •
InfoGa Formo <u>http://</u> ecisior	ain. pre informatio /en.wikipedia <u>n trees</u>	n, go to: org/wiki/Information_gain_in_d

At this stage, the user has access to 3 different methods: InfoGain, OneR and Greedy.



		Super	vised	
	InfoGain			-
	InfoGain			
Neut	OneR			
	Greedy			

InfoGain will operate regularly by reducing the dimensions to the user selected value. OneR operates the same way; however, in this case, the user has a feedback from the algorithm through the text box, giving him the opportunity to check the weights associated to every descriptor.

Phenotypic clustering

For the purpose of this section, let's start buy generating an univariate artificial screening by clicking on *file->Generate Screen->Univariate*



Change the standard deviation of the compound distribution to 0 and push the standard deviation popup button as it is shown on the picture below.



🙅 Ger	erate Screening									
	Number of Plates Columns Rows	10 (m) 24 (m) 16 (m)								
Value distributions	Compounds Distribution Mean Standard Deviation	1000.0 ×								
Bias	Positive Ctrl Column Index Mean Standard Deviation	2 × 1800.0 × 100.0 ×								
	Vegative Ctrl Column Index Mean Standard Deviation	1								
_	Generate									

This means that the compound distribution (Gaussian) will have a variable standard deviation starting 0 and increasing by 30 (defined on the option window as show below) on every plate (here 10 plates are defined).

Import / Export	Display	Clustering	Classification	Correlation matrix								
Misc. Errors	Identif. & C	Correct. Ge	nerate Screenin	g								
Variable parameter steps												
Compour	nd noise St	d Dev.	30.00	×								
	Row effe	ct shift	2.00	×								
	Ratio	x/Y	0.10	×								
	Diffu	sion	1	×								
		0										
		UK										

Click Ok to generate the 10 plates. Starting from:





То



The evolution of the screening data quality can be monitor by clicking on *Screen->Visualization-*>Scatter Points



More precisely, using *Screen->Quality Controls->Descriptor Evolution* and by selecting the class 2 (here the compounds), we obtain:





<u>Note:</u> As only the compounds distribution has been subjected to a variation over the plate index, a quality control such as the Z-factor should not vary over the screening. This can be assess by clicking on Screen->Quality Controls->Z-score



The stacked histogram (available by clicking on *Screen->Visualization->Stacked Histogram*) provide also a convenient way of displaying the data distribution:



Let's jump directly to the clustering step by clicking on the *Classification & Clustering* tab.



KCS analyzer v1.0.3					
File Edit Plate_0	- Plate Scree	en View Plug-ins Help			
Current Plate Dimensionality Redu	ction Systematic Error Identification & Correc	ction Momalization Classification & Clustering Report Export		Class selection Positive (0)	###
Clustering	Classification			Current Descriptor	
K-Means 👻	C4.5 -			Descriptor_0	-
Plate By Plate Q Full Screen	Plate By Plate Full Screen			Descriptor List	
				Descriptor_0	
Cluster Number 2 🚔 Automated	To Be Classified 2				
K-Means.	C4.5.				
http://en.wikipedia.org/wiki/K-means_clustering	http://en.wikipedia.org/wiki/C4.5_algorithm				
Cluster	Classify		anal/zer		

First select the clustering algorithm:

	1.00
HCS analyzer v1.0.3	
File Edit	
Current Plate	duc
Clustering	
K-Means 🖵	
K-Means EM	
(Hierarchical	
Cluster Number 2 🚔 Automated	

When using K-Means, the user has to select the number of expected class. Let's choose 3 (2 controls + 1 compound distributions).

Clustering
K-Means 👻
Plate By Plate
Cluster Number 3 Automated
K-Means. For more information, go to: <u>http://en.wikipedia.org/wiki/K-means_clustering</u> Warning: in such mode the results can be inconsistent from one plate to another.



First, we will start by checking the *Plate By Plate* radio button. In this case, the clustering will be operated on each plate independently.

<u>Note:</u> In such case, no consistency is guarantee for the class association from one plate to another.

Click on Cluster to process the entire screening. Let's come back to the plate visualization tab, and press the popup button *Display Class* to monitor the result.



As said above, the class association is random (here the compound and the controls have been inverted). However, in term of clustering, the process was successful by differentiating the controls and compounds properly.

As the noise is increasing the clustering becomes fuzzy as show on the next figure:



In such case, the Screen->Visualization->Scatter Points give us a good overview of the phenomena.





Let's apply the same clustering method but this time performed in one time over the entire screen.

Clustering	
Clustening	
K-Means	•
Plate By Plate	Full Screen
Cluster Number 3	Automated
K-Means. For more information, go to: http://en.wikipedia.org/wik Waming: in such mode the inconsistent from one plate Plate by plate clustering do	i/K-means clustering results can be to another. ne !

The resulting scatter points graph gives:



As usual the class association is still random, but now as the screening is process globally, there is a consistence from one plate to another.

By choosing another clustering approach such as the Expectation–maximization algorithm (EM), it is possible to automatically identify the optimal number of cluster. Let's perform it his way



Clustering
EM 👻
Plate By Plate Full Screen
Cluster Number 3 💽 Automated
EM. For more information, go to: <u>http://en.wikipedia.org/wiki/Expectation_maximizati</u> on Warning: this task can be time consuming. If the number of class is higher than 10, the clustering will not be performed. Warning: in such mode the results can be inconsistent from one plate to another.

Note: This operation can be time consuming.

The number of identified clusters can be check on the dedicated console.

Clustering	
EM	
Plate By Plate	
Cluster Number 3 💽 Automate	d
<pre>Warning: this task can be time consuming. If the number of class is higher than 10, the clustering will not be performed. Warning: in such mode the results can be inconsistent from one plate to another. Plate_0: 3 cluster(s) Plate_1: 3 cluster(s) Plate_2: 3 cluster(s) Plate_3: 3 cluster(s) Plate_4: 3 cluster(s) Plate_5: 3 cluster(s) Plate_5: 3 cluster(s) Plate_6: 4 cluster(s) Plate_7: 4 cluster(s) Plate_8: 2 cluster(s) Plate_9: 3 cluster(s) Plate_9: 3 cluster(s) Plate_9: 3 cluster(s) Plate by plate EM clustering done !</pre>	A E

It is interesting to see that on plate_6, four clusters have been detected.





In this case, the controls have distinguished as well as two phenotypes within the compounds class.



If the same operation is performed in the entire screening in once, 7 clusters are obtained.

<u>Note:</u> if the cluster number identified by the EM is higher than 10, then the process is cancelled.





For such an approach, the stacked histogram is the most relevant way for understanding the clustering.



The controls have been properly identified; however, the compound distribution has been split in five distinct Gaussian distributions.

Class selection rt ositive (0) Ŧ Positive (0) ۸ 1971.142 Negative (1) 2 Global 3 Ξ 4 5 Global only 6 selected 7

Now we can convert the entire screen wells into class2 (click on Global).

Then un-select the two controls, in order to remove them from further analysis.





The plate design should look like this:

🚣 HC	S an	alyzer v1.0.3																							
File	E	dit Plate_6									- P	late	Screen	Viev	v Plu	ug-ins	Help								
		Current Plate	៸_➡∟	Dimensio	onality R	Reduction	n 🚺	<u>ı</u>	Systema	tic Error	Identific	ation &	Correctio	n [∽	Nomali	ation	14/2	Class	sification	& Cluste	ering	Report Export		Class selection Unselected (-1)
	_	1 2	3 4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	1971.142	Current Descriptor
A	H		_			L		-							-	⊢						L_		Global	Descriptor_0 -
в	H	_	_					-			—			-	-	-									Descriptor List
	-		-			-		-	-						-	-	-					-		Global only selected	Descriptor_0
E	H		-						+		-			-	-	-	-								
F	H					-		-	+-	-	-		-			-	-	-							
G	F					-			-	-			-					-				-		Apply to all plates	
н																									
I.																									
J																								Display class	
K.	-																								
	-		-			-		-	-							-									
N	-		-			-		-	-						-	-						-		19.168	
0	ŀ					-		-	-							⊢									
Р									1							F								S 8	

The idea now is to perform a standardization of the data. For that, click on the *Normalization tab*, select *Standardization* and apply it to the class2 previously defined.



This operation is one of the few process that will modify the data.

Warning	
	By applying this process, data will be definitively modified ! Proceed ?
	Yes <u>N</u> o

Due to the data structure, the first plate has not been processed (null standard deviation).





The full screening scatter points are then difficult to read.



This can be fixed by removing the plate0 from the list of active plate (View-> Plates Manager).



The scatter points become then:



Finally, we can apply an EM clustering with an automated number of cluster identification.





As expected, only one cluster has been identified, as it is display on the histogram visualization.



HCS analyzer provides also a dendogram visualization of the data. The visualization is common in biology and provides a convenient way to check the signature distributions over a limited number of experiments. Let's start by generating a multivariate screening composed of a single plate with a 10 dimensions signature as follow:



H	è Formf	ForMultivariat	teScreen																x
							Dimension	10		Plate Dim Numb	ension er of Plates	10	×						
							Dimension			Co F	lumns lows	24 4	×						
L		Name	Column		Selection	Mean0	Stdv0	Mean1	Stdv1	Mean2	Stdv2	Mean3	Stdv3	Mean4	Stdv4	Mean5	Stdv5	Mean6	Ste
	•	Phenotype 0	0	•	V	0	20	0	20	0	20	0	20	0	20	0	20	0	20
		Phenotype 1	1	•	V	50	20	50	20	50	20	50	20	50	20	50	20	50	20
		Phenotype 2	Entire plate	•	V	100	20	100	20	100	20	100	20	100	20	100	20	100	20
		Phenotype 3	3	•		150	20	150	20	150	20	150	20	150	20	150	20	150	20
		Phenotype 4	4	•		200	20	200	20	200	20	200	20	200	20	200	20	200	20
L		Phenotype 5	5	•		250	20	250	20	250	20	250	20	250	20	250	20	250	20
L		Phenotype 6	6	•		300	20	300	20	300	20	300	20	300	20	300	20	300	20
L		Phenotype 7	7	•		350	20	350	20	350	20	350	20	350	20	350	20	350	20
L		Phenotype 8	8	•		400	20	400	20	400	20	400	20	400	20	400	20	400	20
		Phenotype 9	9	•		450	20	450	20	450	20	450	20	450	20	450	20	450	20
	•																		•
										Generate	•								

Go to the menu Plate->Visualization->Hierarchical Tree

🙅 HCS analyzer v1.0.3			- • • • • •
HCS analyzerv10.3 File Edit Plate_0 Torrer Plate Jacobian Structure A 5 5 7 8 9 10 11 A 5 5 7 8 9 10 11 Common Structure A 5 5 7 8 9 10 11 Common Structure D I 1 2 3 4 5 5 7 8 9 10 11 D I 1 1 2 3 4 5 5 7 8 9 10 11 D I 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Plate Screen View Plug-ins Help Visualization ··· ··· Scatter Points Ctrl + H Quality Controls ·· ··· PCA Ctrl + A 2 13 14 15 16 10 ··· ··· PCA Ctrl + A 10 ··· ··· ··· Ctrl - D 13 14 ··· ··· ··· 13 ··· ··· ··· ··· 14 ··· ··· ··· ··· 13 ··· ··· ··· ··· 14 ··· ··· ··· ··· 13 ··· ··· ··· ··· 14 ··· ··· ··· ··· 13 ··· ··· ··· ··· 14 ··· ··· ··· ··· 15 ··· ··· ··· ··· 16 ··· ··· ··· ··· 17 ··· ··· ··· ··· 18 ··· ··· ··· ··· 19 ··· ··· ··· ···	Dusteing IIII Report Export 2 23 24 Global Global Global orly selected Depley to all plates Depley class 0 87	s selection stave (0)
		0.87	

a window as follow should appear





On the very left side of the dendogram, the user can see the color related to the well classes, followed by the well position. Then, next comes the signature associated to the well.

<u>Note:</u> each descriptor is separately normalized between -1 to +1.

And finally the tree by itself. One can access the tree options by displaying the *option window* and selecting the *Clustering->Hierarchical* box.

tions			
Misc. Errors	dentif. & Correct.	Generate Screenin	g
Import / Export	Display Cluste	ring Classification	Correlation matrix
Hierarchical			
Distan	ce	Euclidean	•
Link T	уре	SINGLE	•
		Ok	

Note: those options will be also applied for the hierarchical clustering process





The user has to remember that such hierarchical clustering computation is highly related to the number of wells involved. Thus, the computational time can increase a lot for high number of data points, and the related dendogram display becomes quickly unclear.

Phenotypic classification

For a clear understanding of the phenotypic classification, we will start this section by generating an artificial multivariate screening.

File	Edit					
6	Load Screen	Ctrl+L				1
	Import Screen	Ctrl+I		Dimensionality Reduction	י 🔔	Sy
	Generate Scree	n	•	Univariate	Ctrl+G	
	Save Screen		►	Multivariate C	trl+Alt+G	
	Add Plates	Ctrl+A				
$\frac{2\pi n}{r}$	Link Data					
5	Exit	Ctrl+X				

Click on *file->Generate Screen->Multivariate*.

Increase the dimension of the data to 4, and keep the original set-up for the other parameters.



		Dime	nsion 4	4	3	Plate Dimension Number of I Column Rows	n Plates s	10 24 16			
Name	Column	-	Selection	Mean0	Stdv0	Mean1	Stdv1	Mean2	Stdv2	Mean3	Stdv3
Phenotype 0	0	•		0	20	0	20	0	20	0	20
Phenotype 1	1	•	V	50	20	50	20	50	20	50	20
Phenotype 2	Entire plate	•	v	100	20	100	20	100	20	100	20
Phenotype 3	3	•		150	20	150	20	150	20	150	20
Phenotype 4	4	•		200	20	200	20	200	20	200	20
Phenotype 5	5	•		250	20	250	20	250	20	250	20
Phenotype 6	6	•		300	20	300	20	300	20	300	20
Phenotype 7	7	•		350	20	350	20	350	20	350	20
Phenotype 8	8	•		400	20	400	20	400	20	400	20
Phenotype 9	9	•		450	20	450	20	450	20	450	20
						Generate					

In such a mode 10 plates containing 3 different phenotypes (represented by multivariate Gaussian distribution, in a 4 dimension space, with means and standard deviations defined in the control). Phenotype 0 and 1 will be located respectively at column 0 and 1.

Click Generate.



You can look at the points distributions along the different axis by clicking on *Screen (or Plate) -> Visualization -> XY scatter points*, and by choosing different descriptors along the axis.





In such example, the phenotypes are clearly defined and well separated. We can also visualize the linear discriminant analysis (LDA) on a plate. To do that, click on Plate->Visualization->LDA



The user has to select a class that will not be taken into account for LDA computation but by projected in the resulting space. Let's choose Phenotype 2 (in dark green).





The horizontal axis of this space represents the linear component that maximize here the separation between the phenotypes 0 and 1 (light green and red).

If we consider those phenotypes as controls, we can then perform a supervised classification by displaying the corresponding tab.

🚣 HCS analyzer v1.0.3		- • • ×
File Edit Plate_0	- Plate Screen View Plug-ins Help	
Current Plate Dimensionality Reduc	tion 🔥 Systematic Error Identification & Correction 🖾 Normalization	Class selection Positive (0) 16
Clustering	Classification	Current Descriptor
K-Means 👻	C4.5 -	Descriptor_0 -
		Descriptor List
Mate by Nate We full Screen Ouster Number Duster Number	W Hate by Hate Full Screen To Be Classified 2 For more information, go to: - http://en.weikoedia.org/weik/C4.5_algorithm	Oecototor_0 Oecototor_1 Oecototor_1 Oecototor_1 Oecototor_2 Oecototor_2 Oecototor_3
	HS	5
Guster	Cessofy	

The classification algorithm can be selected here.

	C4.5		-
	C4.5 Support Vector	or Machine	
0	Neural Netwo K Nearest Nei	rk ighbor(s)	
То	Be Classified	2	



if the *Plate By Plate* radio button is checked, then each plate will be considered independently. It means that, the learning step will be performed on each plate and the classification will follow. If the *Full Screen* radio button is checked then, the learning step will be performed on all the screening data, then applied to all the wells.

<u>Note:</u> In a full screen process, the user will have to perform normalization step to make the data consistent.

The user has to specified a *To Be Classified* class. This class will not be taken into account for the learning step, but will be (as <u>all the other wells</u>) classified. Typically, this represent the screened compounds.



Let's choose Class 2, and then press *Classify* button.

CS analyzer v1.0.3			
Edit Plate_0		en View Plug-ins Help	
Current Plate Dimensionality Rec	duction Systematic Error Identification & Correc	tion Amselization Classification & Clustering Report Export	Class selection Postive (0) - 16
Kfleans Rate By Plate Kfleans Rate By Plate Kfleans Kfleans	Cassification C4.5 C4.5 For more the field of the field o	Process over ! C4.5 classification process finished ! Press (Ctil-T) for current plate tree. OK	Current Descriptor Descriptor_0 Descriptor_0 Descriptor_1 Descriptor_2 Descriptor_2 Descriptor_3
Cluster	Classify	anal/z	er

When done, information about the classification process is displayed, and in the case of C4.5 classification, the user has the opportunity to access each resulting tree. The trees are associated to each plate, and then the user has to selected the plate first and after either using the shortkey *Ctrl+T* or click on *Plate->Visualization->Classification Tree*.





By going back to the plate visualization, the user can assess the quality of the results.



In this specific case, the classification is obvious, and the compounds have been classified in the same category as the phenotype 1.





siRNA screening

In the context of siRNA screening, and if the user has locus ID associated to each well, we offer the possibility to connect the application to gather information about the genes involved as well as the pathways.

While loadint the screening the user has to specify the Locus ID such as follow:

Impo	rt								X
Plate	Dimensions								
Co	lumns 24								
Ro	ows 16	×							
	Data Name	Selection	Туре		Readout 0	Readout	1	Readout 2	-
	Plate384		Plate name	•	SRP000459	SRP0004	79	SRP000502	
	Well 384w		Well position	-	D24	P20		G11	
	Cat		Descriptor	-	Druggable	Druggable	•	Druggable	
	Catalog Number		Descriptor	-	1	2		3	
	Gene Symbol		Descriptor	-	1	2		3	-
	Description		Descriptor	-	adrenocortical dy	deleted in	esoph	hypothetical pro	o'
0	Locus ID		Locus ID	•	65057	115123		57574	
	Fluo1		Descriptor	-	0.8228594	0.723350	4	0.9090863	
	Size		Descriptor	-	0.9823682	0.987038	5	0.7240621	-
	Volume		Descriptor	-	59.26911984	183.9027	118	137.0042838	
	Ratio 1		Descriptor	-	144.2518279	234.0326	818	139.0108669	
	Fluo2		Descriptor	-	90.0408859	176.5919	403	74.81596763	1
•	Tosturo		Descriptor	1_	104 5004000	221 5572	200	105 0001450	-
								Ok	

In such context, the locus ID will be associated to each well. By right clicking and selecting *Info* from the contextual menu, the user can check the locus ID of a specific well.





The user can also select Kegg to initiate a connection to the Kegg server.



Note: for the option an internet connection is required.

If information about the gene are available, a window such as follow will be displayed.



Where the gene of interest is highlighted in orange. In parallel, another window displays the pathway(s) related to this gene.





By double clicking on the pathway of interest, another window pops up, providing information about it.



In order to extract automatically information about the pathways involved in a certain phenotype, the user can select the following menu *Screen->Gene Analysis->Pathways analysis*.





The application will then parse all the genes/well of the current selected class and gather the pathways related. At the end of the process, a pie chart displays the ratio of pathways involved and extracts the most recurrent.



<u>Notes:</u> Due to the internet connections, this process is highly time consuming. This operation can be performed during the report process.



Development

Start by extracting the complete solution from the "HCSAnalyzer.zip" file.



Double click on the solution file to launch the development environment tool.

rganize	▪ ∞ Open ▪ Burn New folde	r				BE	• 🗊	
*	Name	Date modified	Туре	Size				
	🅌 bin	2/20/2012 10:16 AM	File folder					
	🕌 Classes	2/20/2012 10:16 AM	File folder					
	퉬 Controls	2/20/2012 10:16 AM	File folder					
	🐌 Forms	2/20/2012 10:16 AM	File folder					
	📕 HCSPlugin	2/20/2012 10:16 AM	File folder					
	퉬 obj	2/20/2012 10:16 AM	File folder					
	Properties	2/20/2012 10:16 AM	File folder					
	Resources	2/20/2012 10:16 AM	File folder					
	Service References	10/18/2011 2:12 PM	File folder					
	Web References	2/20/2012 10:16 AM	File folder					
	app.config	2/20/2012 9:29 AM	CONFIG File	2 KB				
	Class1.cs	10/4/2011 3:37 PM	Visual C# Source f	1 KB				
	Class2.cs	2/20/2012 9:29 AM	Visual C# Source f	1 KB				
	Classif.cs	2/20/2012 9:27 AM	Visual C# Source f	20 KB				
	Clustering.cs	2/20/2012 9:29 AM	Visual C# Source f	18 KB				
	DimensionalityReduction.cs	2/20/2012 9:29 AM	Visual C# Source f	16 KB				
	HCS Analyzer.csproj	2/20/2012 10:07 AM	Visual C# Project f	33 KB				
	≳ HCS Analyzer.csproj.user	2/20/2012 9:22 AM	Visual Studio Proj	2 KB				
-	😕 HCS Analyzer.sln	2/20/2012 10:07 AM	Microsoft Visual S	3 KB				
C	Import_Export.cs	2/20/2012 9:29 AM	Visual C# Source f	95 KB				
-	Mormalization.cs	2/20/2012 9:29 AM	Visual C# Source f	13 KB				
ā.,	SystematicErrorCorrection.cs	2/20/2012 9:29 AM	Visual C# Source f	6 KB				
	WindowHCSAnalyzer.cs	2/20/2012 10:01 AM	Visual C# Source f	165 KB				
	WindowHCSAnalyzer.Designer.cs	2/20/2012 9:29 AM	Visual C# Source f	152 KB				
	WindowHCSAnalyzer.resx	2/16/2012 9:46 AM	RESX File	162 KB				
a H	WindowHCSAnalyzer.resx	2/16/2012 9:46 AM	RESX File	162 KB				
18 18 -								

One solution ("HCS Analyzer") containing two projects ("HCS Analyzer" and "HCSPlugin") has to be loaded.





Press debug to launch the software.

HCS Analyzer (Running) - Microsoft Visual Studio	(Administrator)	
File Edit View Project Build Debug Team	Data Tools Architecture Test Analyze Window Help	
i 🗊 • 🖽 • 🚰 🛃 📓 🕺 🕹 📇 🔊 • @	· [J · G] > Debug · @ Listells · 직원 및 전상 환원 전 · 디 김 일 사 昭 译류 [그 일 그 위식 위식 원 용	🙊 📮 E 🕨 💷 🚆 👘
Toolbox 👻 🕂 🗙	WindowHCSAnalyzer.cs 🖲 🗙 🗸 Solution Explorer	→ # ×
▲ General There are no workly controls in this move. Deep	• HCSAnalyzer HCSAnalyzer → • • •	
an item onto this text to add it to the toolbox.	Image: Constraint of the second se	E
HCS analyzer v1.0.3	getter (getterne internet) - getter	- • ×
File Edit	 Plate Screen View Plug-ins Help 	
Current Plate	Dmensionality Reduction 🚺 Systematic Error Identification & Correction 🚺 🗠 Normalization 🔛 Classification & Clustering	Class selection Positive (0) - #####
Ready	HEF Gobal Gibal rhy sected Poply to al places Deplay clase Deplay clase Line LivyZeyr Rine LivyZeyr	Current Descriptor Descriptor List



Plugin Development

Open the HCSAnalyzer solution as describe above. Add a new project to the solution.



Choose the "HCSAnalyzer Plugin" template and specify the name:



Add New Project							? ×
Recent Templates		.NET Fra	amework 4 🔹 Sort by: Default			Search Installed Templates	٩
Installed Templates Visual C#		1	ASP.NET MVC 2 Web Application	Visual C#	*	Type: Visual C# <no available="" description=""></no>	
Windows Web		¢C#	Silverlight Application	Visual C#			
 Office Cloud Reporting 		c ≇	Silverlight Class Library	Visual C#			
 SharePoint Silverlight 		C#	WCF Service Application	Visual C#			
Test WCF		1	ASP.NET Dynamic Data Entities Web	ApplicationVisual C#			
Workflow ▷ Other Languages			Enable Windows Azure Tools	Visual C#			
 Other Project Type Database Modeling Designets 	25	C#	Excel 2010 Workbook	Visual C#	E		
 Test Projects 		<mark>⊘c</mark> #	Outlook 2010 Add-in	Visual C#			
Online Templates		C#]	Word 2010 Document	Visual C#			
		*	Activity Library	Visual C#			
		c#	WCF Workflow Service Application	Visual C#			
		anatyzar	HCSAnalyzer Plugin	Visual C#			
<u>N</u> ame:	My HCS Analyze	er PlugIn					
Location:	C:\DevForTemp	late\			•	Browse	
						ОК	Cancel

<u>Note:</u> To have access to this template, the "HCSAnalyzer Plugin.zip" has to be present in your Visual Studio template directory (e.g.:

"C:\Users\Myself\Documents\Visual Studio 2010\Templates\ProjectTemplates").

At this stage you need to add two references to your plugin project :

· · · · · · · · · · · · · · · · · · ·	Solution Explorer Image: Constraint of the second seco

Click on the "Browse" tab, and go to your HCSAnalyzer development directory (probably something like "C:\HCSAnalyzer\bin\Debug") and select "HCSAnalyzer.exe" and "HCSPlugin.dll"



Look in: 🌗 Debug	- 🕝 🏂 📂 🛄 -			
Name	Date modified	Туре	Size	4
\mu Plugins	2/20/2012 3:06 PM	File folder		
Accord.dll	1/19/2012 11:25 AM	Application extens	6 KB	
🚳 Accord.Math.dll	1/3/2012 10:35 PM	Application extens	231 KB	
Accord.Statistics.dll	1/3/2012 10:36 PM	Application extens	246 KB	
🚳 AForge.dll	1/19/2012 11:28 AM	Application extens	17 KB	1
🚳 AForge.Math.dll	12/12/2011 8:18 PM	Application extens	67 KB	
🚳 AlgLib.dll	4/27/2011 12:34 PM	Application extens	927 KB	
验 HCSAnalyzer.exe	2/20/2012 2:00 PM	Application	2,831 KB	
HCSAnalyzer.exe.manifest	2/20/2012 2:00 PM	MANIFEST File	30 KB	
HCSAnalyzer.vshost.exe	2/20/2012 3:26 PM	Application	12 KB	
HCSAnalyzer.vshost.exe.manifest	2/20/2012 2:00 PM	MANIFEST File	30 KB	
N HCSPlugin.dll	2/20/2012 2:00 PM	Application extens	10 KB	
🚳 ICSharpCode.SharpZipLib.dll	9/10/2010 11:24 AM	Application extens	196 KB	
IKVM.AWT.WinForms.dll	3/17/2011 6:31 AM	Application extens	176 KB	
IKVM.OpenJDK.Beans.dll	3/17/2011 6:31 AM	Application extens	221 KB	
IKVM.OpenJDK.Charsets.dll	3/17/2011 6:31 AM	Application extens	4,853 KB	
IKVM.OpenJDK.Corba.dll	3/17/2011 6:31 AM	Application extens	2,144 KB	
IKVM.OpenJDK.Core.dll	3/17/2011 6:31 AM	Application extens	3,155 KB	
IKVM.OpenJDK.Jdbc.dll	3/17/2011 6:31 AM	Application extens	386 KB	
IKVM.OpenJDK.Management.dll	3/17/2011 6:31 AM	Application extens	1,324 KB	
IKVM.OpenJDK.Media.dll	3/17/2011 6:31 AM	Application extens	813 KB	
File name: "HCSPlugin.dll" "HCSAnalyzer.exe"				•
Files of type: Component Files (* dli* tlb * olb * ocr:	eve:*manifest)			

Once compiled, your plugin needs to be located in your HCSAnalyzer plugins directory. To do that automatically, double click on plugin Properties, and go the "Build Events" tab:

Visual Studio (Administrator)	
Build Debug Team Data Tools Architecture Test Analyze Window Help	
X My HCS Analyzer Plugin* ×	Solution Explorer T X
Application Configuration: N/A Platform: N/A	Solution 'HCS Analyzer' (3 projects) Get HCS Analyzer'
Build Pre-build event command line:	HCSPlugin HCSPlugin
Debug	Properties Construction AssemblyInfo.cs Besources.resx
Resources	 Settings.settings References
Services Edit Pre-build	- HCSAnalyzer - HCSPlugin
Reference Paths	System System.Core
Signing *	- System.Windows.Forms
Edit Post-build	

and specify your plugins directory in the "Post-build event command line:" (here "copy "\$(TargetPath)" "E:\HCSAnalyzer\bin\Debug\Plugins" ")



HCS Analyzer - Microsoft Visi	ual Studio (Administrator)	
<u>File Edit View Project Bu</u>	ild <u>D</u> ebug Tea <u>m</u> D <u>a</u> ta <u>I</u> ools Ar <u>c</u> hitecture Te <u>s</u> t A <u>n</u> alyze <u>W</u> indow <u>H</u> elp	
i 🛅 • 🖽 • 💕 🛃 🔊 🐰	🔄 🔠 🔊 • 🗠 - 💭 • 🖏 🕨 Debug 🔹 🔯 🌆 🚺 🔹 • 🖓 🕾 🖓 🖄 🛠 🛃 🖬 • 🖕	
Toolbox 🝷 🕂 🗙	My HCS Analyzer Plugin* 🗙	← Solution Explorer ← 및 ×
A General There are no usable controls in this group. Drig an illum onto this text to add it to the toolbor.	Application Build Configuration: N/A Platform: N/A Build Events* Pre-build event command line: Image: Configuration: Image: Configuration	 Solution HCS Analyzer (3 projects) HCS Analyzer (3 projects) HCS Nalyzer HCSPlugin Propertiesi AssemblyInfo.cs Resourcescreax Settings settings References HCSPlugin System.Ore System.Core <l< th=""></l<>
I I		
		- # ×
🔆 Toolbox 💐 Server Ex	👷 Error List 💷 Output 🛒 Find Results 1 🇯 Find Symbol Results 🛼 Pending Changes 🔠 Code Metrics Results	
Creating project 'My HCS Analyz	er Plugin' project creation successful.	

Finally give a name and a menu location to your plugin by clicking on the "Resources" tab and by filling out the three corresponding cells:

👓 HCS Analyzer - Microsoft Vis	sual Studio (Administrato	or)				_ D _ X
File Edit View Project B	uild Debug Team I	Data Tools Archit	ecture Test Analyze Window Help			
i 🔄 - 🖦 - 😝 📮 🎯 🗐	B (B) 9 - (1 -)	🖂 - 📖 🕨 Debi	10 - 100 IM - 1	N 🕾 🖓 🕸 🛠 🖬 🕅 🖬 + 🗌		
Toolbox ♥ Ψ X	My HCS Analyzer Plugi	a* ×			Solution Explorer	• # X
There are no usable controls in this group.	Application Build	ate Strings 👻 🎦 Ac	ld <u>R</u> esource 👻 Re <u>m</u> ove Resource 🛛 📰 × 🗍 Access N	lodifier: Internal 💌	Solution 'HCS Analyzer	/zer' (3 projects)
text to add it to the		Name	 Value 	Comment	My HCS Analyze	er Plugin
toolbox.	Build Events*	author	MySelf		Properties	
	Debug	menu	Data Analysis		(I) Assembly	yInfo.cs
	Resources*	▶ name	Compute EC50		Resource	.s.resx settings
	Services	*			References HCSAppl	brzer
					- HCSPlug	jin
	Settings				- System	
	Reference Paths				- System.C	Lore
					- System.D	Jrawing Mindows Former
	Signing				Form1.cs	Vindowstronnis
	Code Analysis					
					Solution Explorer	a leam Explorer
	1				Properties	- ∓ ×
	Output				• # ×	-
	Show output from:		- 3 4 - - - - - - -	-		
				_	(Name)	name ^
					Comment	=
					Filename	Fach added in service
					Type	System String, mscorlib.
					(Name)	Contraction of the second
					Name used to identify th	te resource in code.
🔆 Toolbox 💐 Server Ex	📸 Error List 📃 Outp	out 🖳 Find Results :	L 🔉 Find Symbol Results 🛛 🕂 Pending Changes 🗮 🤇	Code Metrics Results		
Creating project 'My HCS Analy	zer Plugin' project crea	tion successful.				

You can then develop your plugin and launch the HCSAnalyzer application. Your plugin will be available through the menu "Plug-ins"



