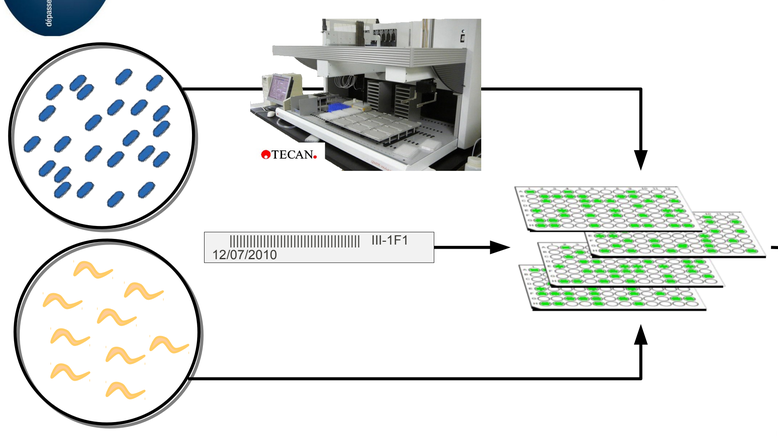
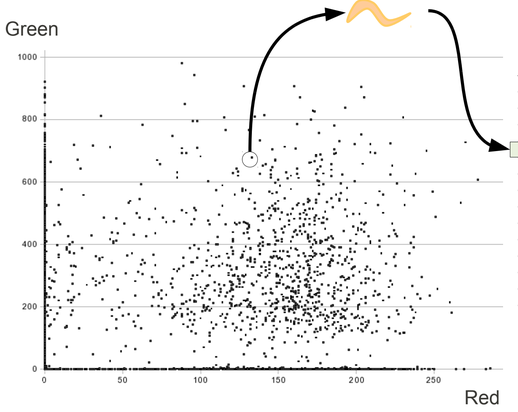
Optimizing analysis of high-throughput screens with the UBI Biosort

Frédéric MONTANANA[1], Jérôme BELOUGNE[1], Aurélie AUTRET[1], Barbara SQUIBAN[1], C. Leopold KURZ[1], Olivier ZUGASTI[1], Renaud JULIEN[2], Fabien CAVASINO[2], Jonathan EWBANK[1]

[1] Centre d'Immunologie de Marseille Luminy, Case 906, 13288 Marseille Cedex 9, France, [2]Modul-Bio, Parc Scientifique Luminy Biotech II, Case 935,13288 Marseille Cedex 9





Id	Plate	Row	Column	Clog	Scan rate	Status sort	Status sel	TOF	EXT	Green	Yellow	Red	PH Ext	PW Ext	PC Ext	PH Green	PW Green	PCGreen	PH Yellow	PW Y
0	0	A	1	N	2500	0	40	144	95	0	0	0	37076	143	n/a	300	142	0	52	9
1	0	A	1	N	2500	0	40	807	776	132	0	144	59336	805	n/a	12944	805	0	52	52
2	0	A	1	N	2500	0	5	596	520	399	0	181	46704	595	n/a	61444	595	5	60	40
3	0	A	1	N	2500	0	40	131	70	0	0	0	32020	129	n/a	488	124	0	48	7
4	0	A	1	N	2500	0	40	129	96	0	0	0	51012	127	n/a	428	124	0	64	7
5	0	A	1	N	2500	0	40	89	37	0	0	0	21792	87	n/a	204	87	0	52	5
6	0	A	1	N	2500	0	5	548	440	732	0	119	40856	547	n/a	65532	547	4	60	37
7	0	A	1	N	2500	0	40	165	164	1	0	0	57692	163	n/a	608	163	0	44	10
8	0	A	1	N	2500	0	40	132	80	0	0	0	30588	130	n/a	260	130	0	52	7
9	0	A	1	N	2500	0	40	158	105	0	0	0	33640	156	n/a	288	156	0	64	10
10	0	A	1	N	2500	0	40	626	657	159	0	170	54164	625	n/a	20584	625	1	48	40
11	0	A	1	N	2500	0	40	121	75	0	0	0	40132	119	n/a	400	118	0	56	8
12	0	A	1	N	2500	0	40	122	82	0	0	0	35560	120	n/a	356	120	0	64	8
13	0	A	1	N	2500	0	40	74	29	0	0	0	18852	72	n/a	200	71	0	48	4

Our laboratory routinely uses worms carrying fluorescent reporter constructs to delineate innate immune regulatory networks. We are performing RNAi screens using the Ahringer and the ORFeome RNAi libraries in order to understand globally the regulation of anti-fungal peptide gene expression.

These quantitative assays are largely automated and performed using a platform built around the UBI Biosort, coupled to a TECAN robot and a bar-code printer to permit a traceability for plates. The Biosort of Union Biometrica allows us to analyze 96-well plates. We are able to have 80-100 synchronized individuals per well.

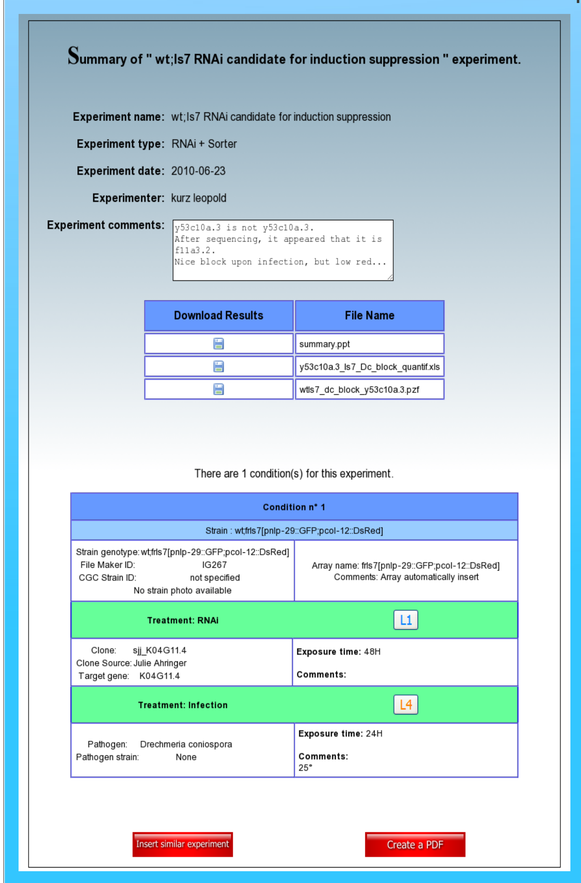
For each worm analyzed in a well, we obtained 17 values. Thus, a large amount of data and metadata are obtained per well and per worm. Finally, replicates are performed and have to be compared. We have developed tools and strategies to maximize the efficiency of the platform, and facilitate data storage and analysis. This involves three parts : metadata storage, data storage and data analysis.

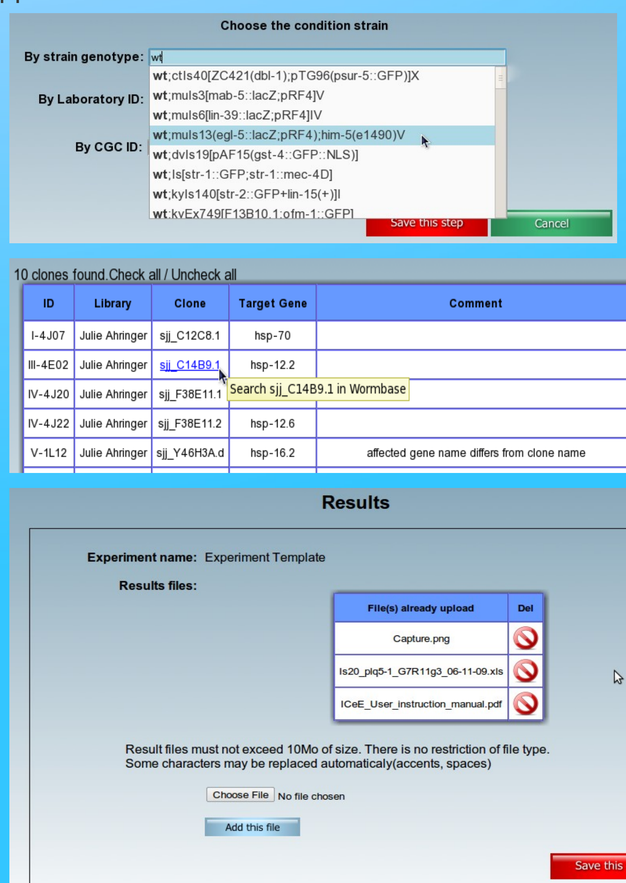
How to store efficiently high throughput method data to be able to retrieve them easily to allow a global interpretation of those data ?

ICeE : e-lab book for metadata storage and data centralization

Interface for Caenorhabditis elegans Experiments — ICeE

1) Easy storage of experimental conditions :
- definition of strain used and treatments applied



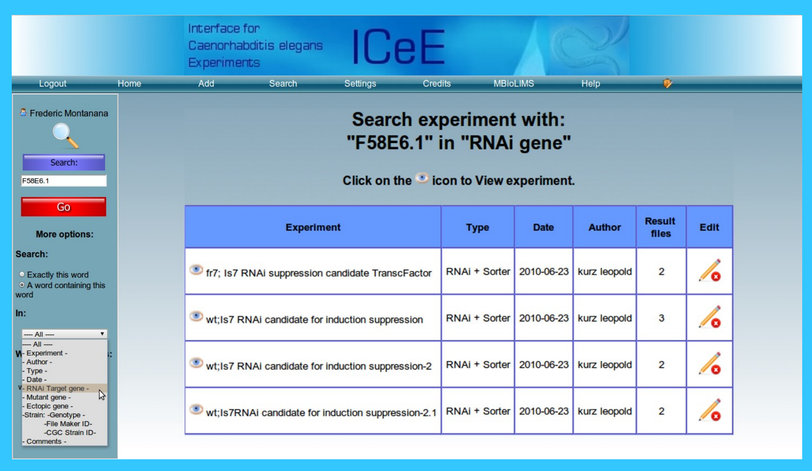


2) CGC strain, ORFeome and Ahringer libraries preloaded and able to centralize all home made databases.

3) Linked to Wormbase to complete info

4) File storage for results files or protocol files. There is no type restriction : (text file, image, video, pdf, tab-delimited file like excel, ...)

ICeE is a web based interface with a database that permit the storage of the condition of the experiment realisation. To be able to analyze an experiment those informations are very important. ICeE is compatible Windows, Mac and Linux and have been tested with the most used web-browsers : Internet explorer (7 and >), Firefox, Safari, Google Chrome and Opera.



5) Search function to make easy the retrieve of an experiment.

6) OpenSource (free to download, modifications allowed and encouraged) and hosted on sourceforge (Wiki, forum, bugs tracker, ...)

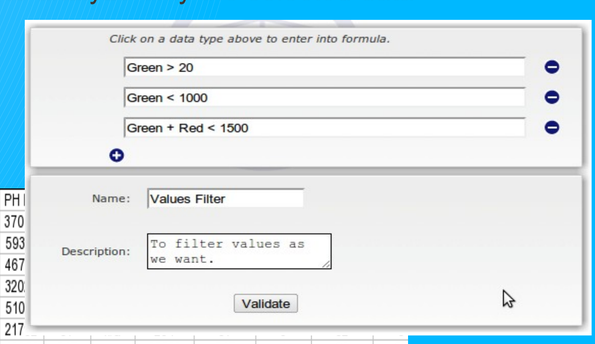
MBioLIMS : LIMS for Biosorter data storage and global analysis

Modul-Bio® High throughput solutions

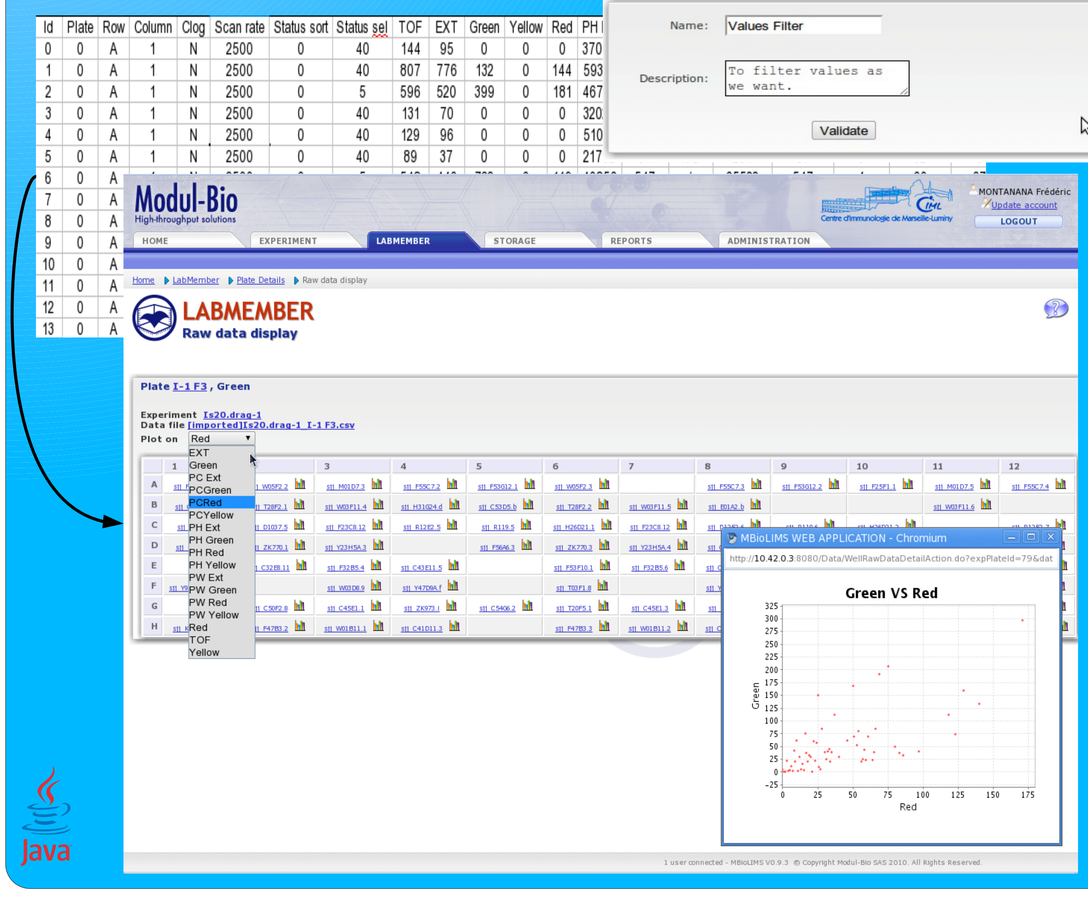
ModulBio had developed a LIMS to pre-treat and analyze biosorter data. The text file generated by the sorter is sent to this database. The results are linked with the metadata uploaded in ICeE.

7) Pre-treatment : sort by any type of value to analyze only the selected data

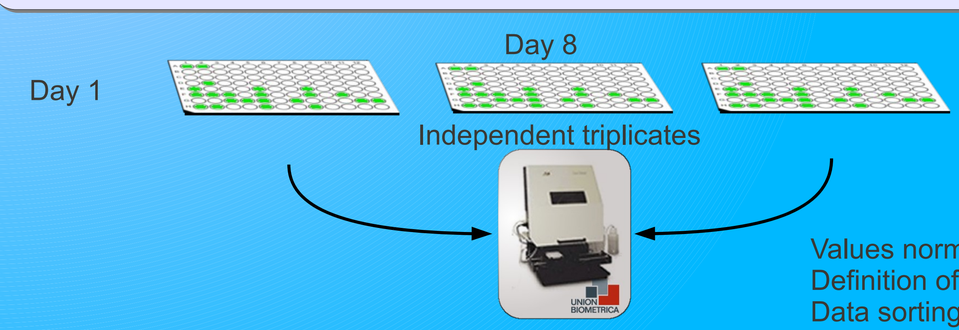
8) Global analysis : automatically generate dot-plots for any selected values from a well.



Id	Plate	Row	Column	Clog	Scan rate	Status sort	Status sel	TOF	EXT	Green	Yellow	Red	PH
0	0	A	1	N	2500	0	40	144	95	0	0	0	370
1	0	A	1	N	2500	0	40	807	776	132	0	144	593
2	0	A	1	N	2500	0	5	596	520	399	0	181	467
3	0	A	1	N	2500	0	40	131	70	0	0	0	320
4	0	A	1	N	2500	0	40	129	96	0	0	0	510
5	0	A	1	N	2500	0	40	89	37	0	0	0	217
6	0	A											
7	0	A											
8	0	A											
9	0	A											
10	0	A											
11	0	A											
12	0	A											
13	0	A											



Specific statistic tools to draw conclusions from complex dataset



Values normalization
Definition of a threshold
Data sorting

ChV6-2_110610
avec tous les individus (saturés et non saturés)

		Green/TOF*100	< 0,6	< 0,8
		Green/Red*100	< 0,6	< 0,8
		D.O. < 1000 dans	1 plaques	2 plaques	3 plaques

Plaque V-6 F1		1	2	3	4	5	6	7	8	9	10	11	12
A	gene	F38E1.3	C24B5.3	ZC196.8	T06E4.6	F38E1.5	C24B5.4	ZC196.9	T06E4.7	F38E1.6	K04	B0507.1	T06E4.8
	Green/TOF*100	1.55	1.29	0.67	2.18	1.83	1.87	1.29	1.87	2.15	1.24	1.56	1.65
	Green/Red*100	1.59	1.10	0.60	1.72	1.44	1.36	1.09	1.44	1.52	1.01	1.26	1.47
B	gene	F52E1.5	F07C3.4	C01B7.3	ZK287.5	F52E1.6	GFP	C01B7.4	ZK287.6	sta-2(CV6_3	F07C3.7	C01B7.5	ZK287.7
	Green/TOF*100	1.19	1.37	1.12	1.42	1.05	0.11	0.64	0.99	0.44	1.22	0.88	2.24
	Green/Red*100	1.20	1.00	1.13	1.48	1.25	0.16	0.64	0.81	0.45	1.21	0.93	1.99
C	gene	T23B12.5	sta-2	H14N18.3	F40F9.9	OP50	F44A2.2	H14N18.4	F40F9.10	T23B12.7	F44A2.3	T19A5.1	snf12
	Green/TOF*100	1.37	0.41	1.22	0.88	1.63	1.11	0.90	1.10	1.78	0.85	0.76	1.52
	Green/Red*100	1.53	0.43	1.20	0.93	1.00	1.10	0.97	1.05	1.60	0.87	0.87	1.39
D	gene	pkc-3	plc-3	F59E11.6	F46B6.2	F45F2.2	K07B1.7	F59E11.7	F46B6.3	F45F2.3	rab-1	F59E11.8	F46B6.4
	Green/TOF*100	1.09	1.20	1.29	1.34	0.63	1.26	1.25	2.44	1.14	0.37	0.94	2.21
	Green/Red*100	1.15	1.22	1.01	1.20	0.56	1.20	1.14	1.69	1.04	0.31	0.86	2.05
E	gene	rack-1	E02C12.11	R03H4.3	C50C10.3	F25G6.2	Ø	R03H4.4	C50C10.4	F25G6.3	F36D4.1	rack-1	Ø
	Green/TOF*100	0.35	0.75	0.80	0.76	0.63	0.85	1.31	1.22	1.38	1.52	0.38	2.45
	Green/Red*100	0.72	0.67	0.81	0.96	0.46	1.56	1.37	1.17	1.06	1.25	0.76	1.68
F	gene	C10B5.3	Ø	C25E10.6	T11A5.2	F41E6.1	K06A4.2	Ø	T11A5.3	F41E6.2	K06A4.3	C25E10.8	Ø
	Green/TOF*100	0.81	1.07	1.49	1.16	2.03	1.28	1.86	1.68	1.39	0.60	0.90	2.03
	Green/Red*100	1.18	1.28	1.15	1.29	1.52	1.28	1.35	1.50	1.38	0.74	0.66	1.80
G	gene	Ø	C50F4.5	T27E4.7	T07C12.1	F41E6.13	C50F4.6	T27E4.8	T07C12.2	F41E6.14	C50F4.7	Ø	T07C12.3
	Green/TOF*100	1.26	1.29	1.51	0.87	1.37	1.13	1.50	2.14	1.60	1.24	1.70	1.84
	Green/Red*100	1.08	1.52	1.49	1.00	0.97	1.24	1.27	1.65	1.28	1.32	1.04	1.90
H	gene	ZC190.7	Ø	f21C10.10	C27H6.3	Ø	F25B3.4	B0222.1	$ C27H6.4	ZC190.9	F25B3.5	B0222.2	K08H10.1
	Green/TOF*100	0.47	0.77	0.92	0.78	1.67	1.04	1.33	0.96	0.74	0.84	0.50	0.99
	Green/Red*100	0.75	0.92	1.20	0.91	1.23	1.08	1.55	0.92	0.88	1.14	0.40	1.32

saturation	
Plq4-RNAiG560R650-CV-6-2-1-11-06-2010.txt	
$GREEN	[1] "5 %"
$nb_GREEN	[1] 236
$GREEN_per	[1] "8 %"
$nb_GREEN_per	[1] 339
$RED	[1] "0 %"
$nb_RED	[1] 0
$RED_per	[1] "0 %"
$nb_RED_per	[1] 0
$nb_total_individus	[1] 4151
Plq5-RNAiG560R650-CV-6-2-2-11-06-2010.txt	
$GREEN	[1] "4 %"
$nb_GREEN	[1] 306
$GREEN_per	[1] "7 %"
$nb_GREEN_per	[1] 489
$RED	[1] "0 %"
$nb_RED	[1] 0
$RED_per	[1] "0 %"
$nb_RED_per	[1] 3
$nb_total_individus	[1] 6985
Plq6-RNAiG560R650-CV-6-2-3-11-06-2010.txt	
$GREEN	[1] "3 %"
$nb_GREEN	[1] 171
$GREEN_per	[1] "4 %"
$nb_GREEN_per	[1] 261
$RED	[1] "0 %"

Conclusion

This kind of architecture allows us to fully exploit results from the Biosort. The plates are traced thanks to the Zebra bar-code printer and every experimental conditions are stored in ICeE to permit an efficient storage and retrieval of the data. The raw data are automatically sent by the Biosort computer to the MbioLIMS database and can be directly analyzed. Finally, despite the complexity of the datasets we obtain, the specific statistical tools we have developed according to our project needs, allow us to draw conclusions.