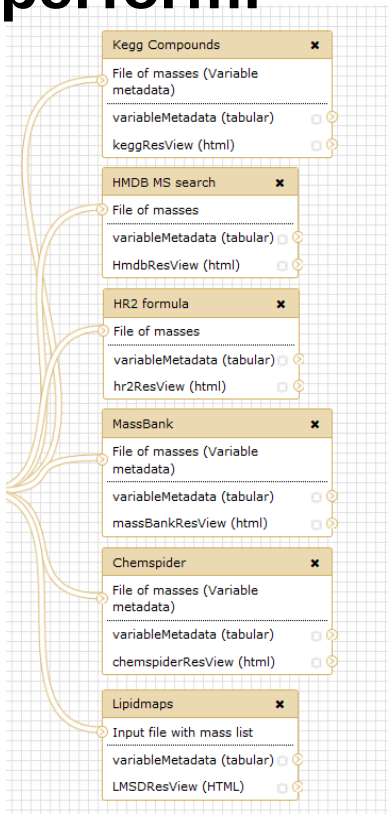
HOW TO PERFORM LC-MS ANNOTATIONS?

W4M Core Team

« LC-MS ANNOTATION » MODULES IN W4M

The « LCMS Annotation » modules allow you to perform:

- « Adducts » annotation
- « Chemical » annotation:
  - ✓ Meta-engine (Chemspider)
  - ✓ General banks (Kegg, ChEBI, PubChem…)
  - ✓ Spectral banks (MassBank)
  - ✓ Specialized banks (HMDB, Lipidmaps, …)
  - ✓ « De novo » tools (HR2 for chemical composition)
- Automatic probabilistic lcms-based metabolome annotation (ProbMetab)



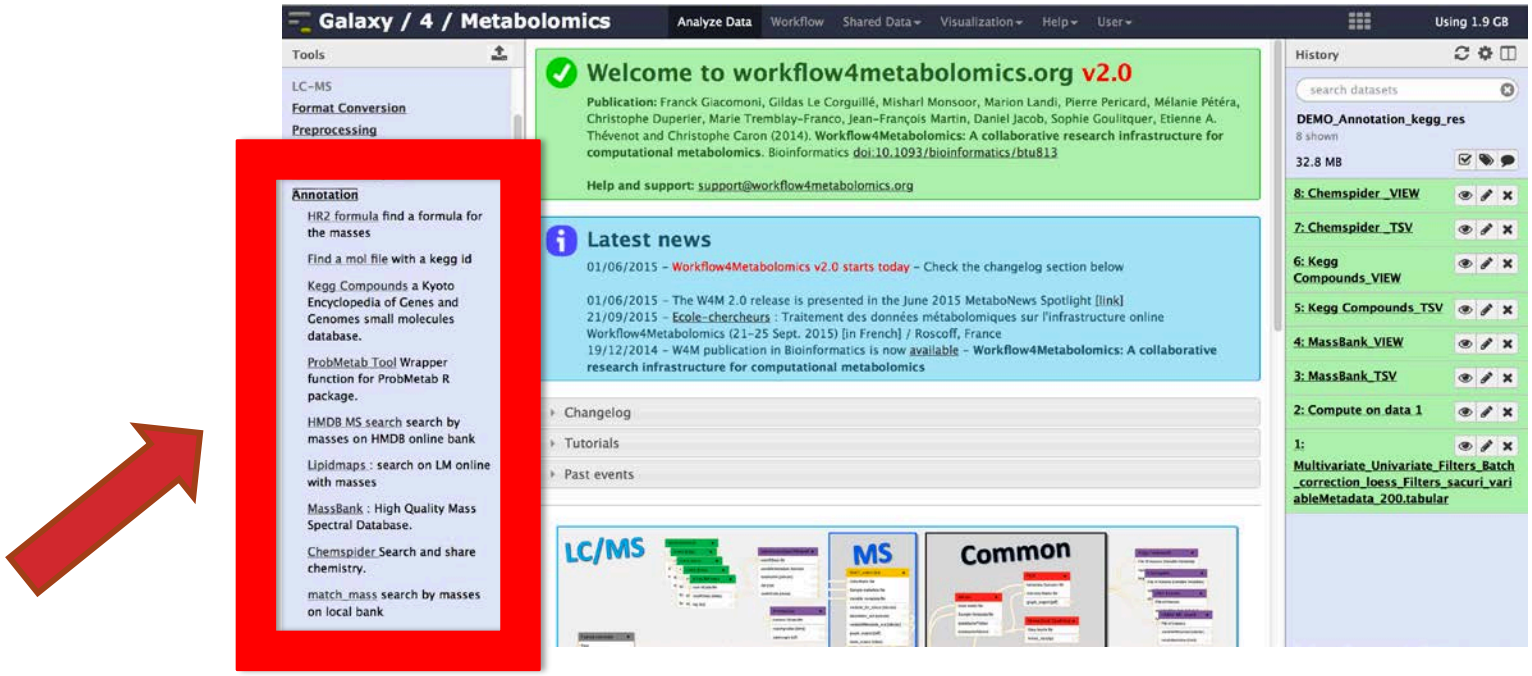
« LCMS ANNOTATION » MODULES IN W4M

Selection of the modules:

- LCMS annotation modules are accessible via the left panel:

> LC-MS

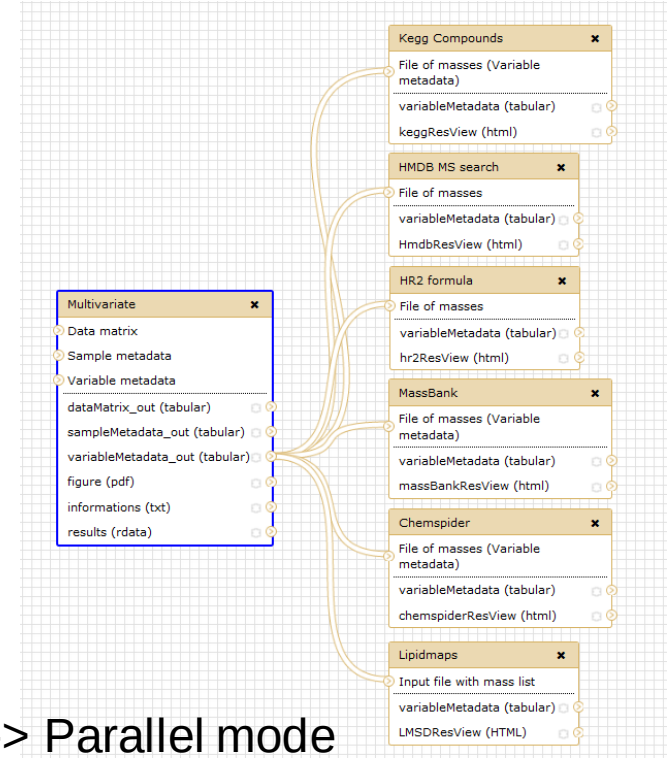
> 6-Annotation



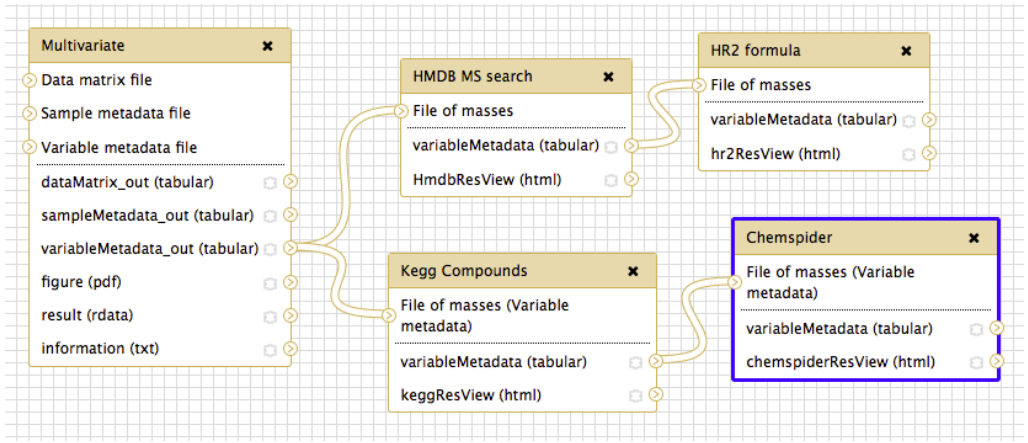
CHAINING THE « ANNOTATION » MODULES

The « LCMS Annotation » modules can be chained with « statistical » modules and the Filters module. ProbMetab can be directly connected with XCMS outputs modules.

Query modules can be run sequentially or in parallel.



-> Parallel mode



-> Sequential mode

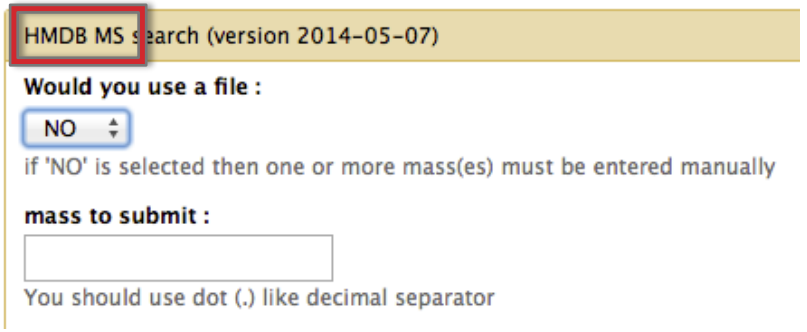
YOUR INPUT DATA

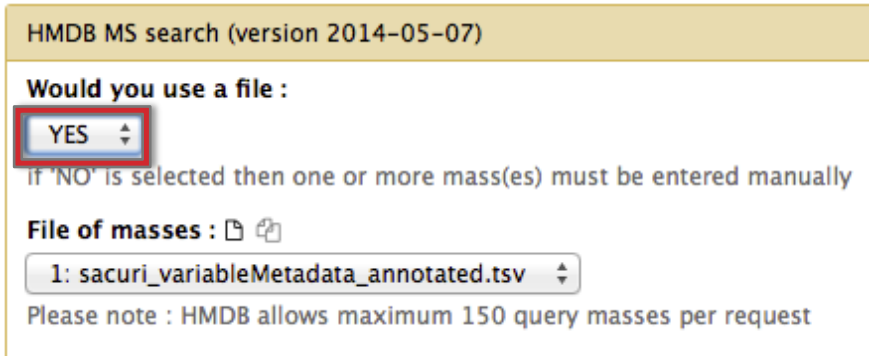
« Chemical Annotation » modules take as input either:

- a list of masses entered manually

or

- an input file





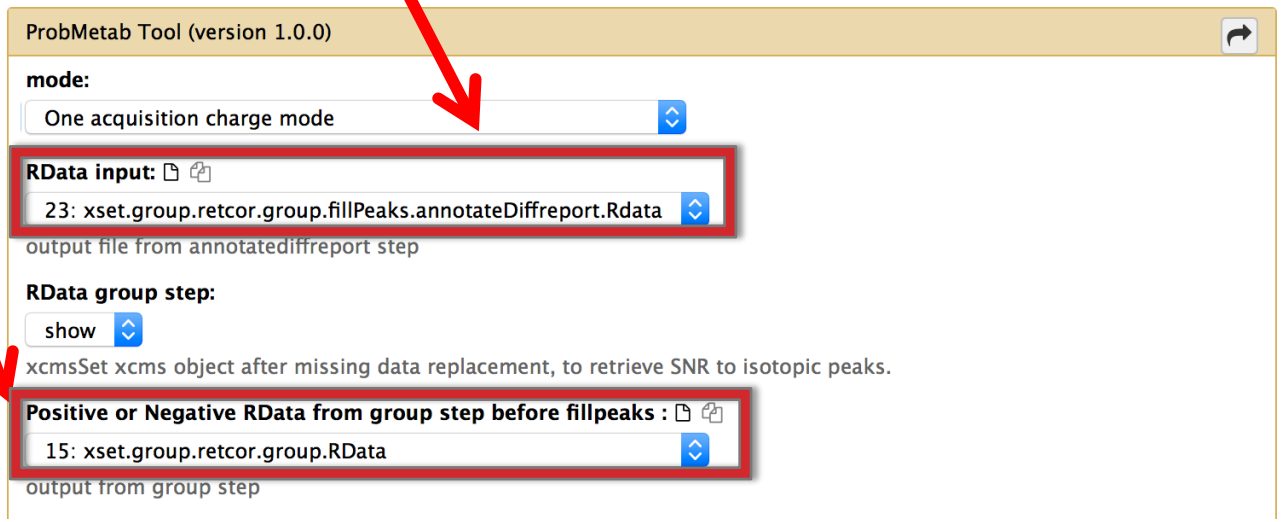
YOUR INPUT DATA

« ProbMetab » module take as input either:

- output file from annotatediffreport

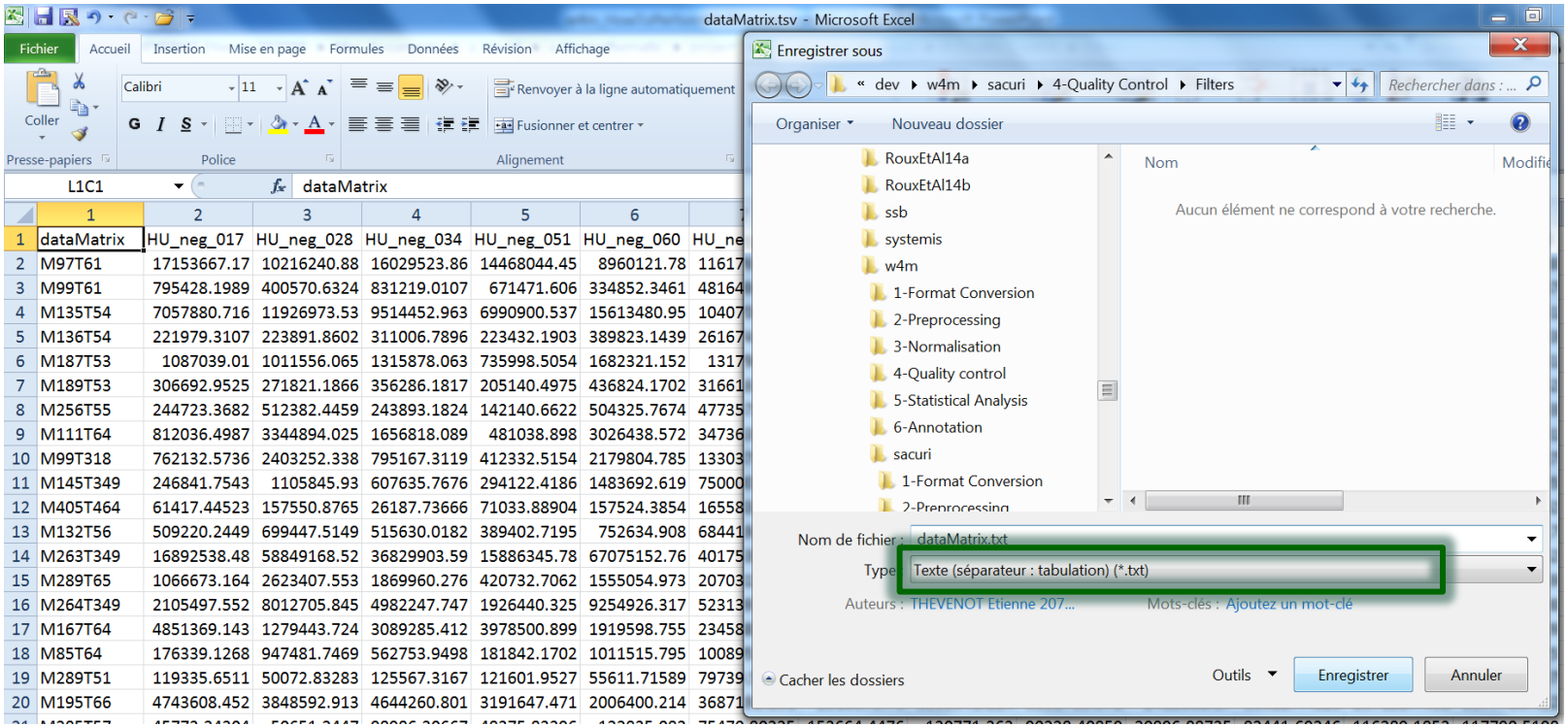
and

- xcmsSet xcms object after missing data replacement, to retrieve SNR to isotopic peaks.



PREPARING YOUR INPUT FILE (1/3)

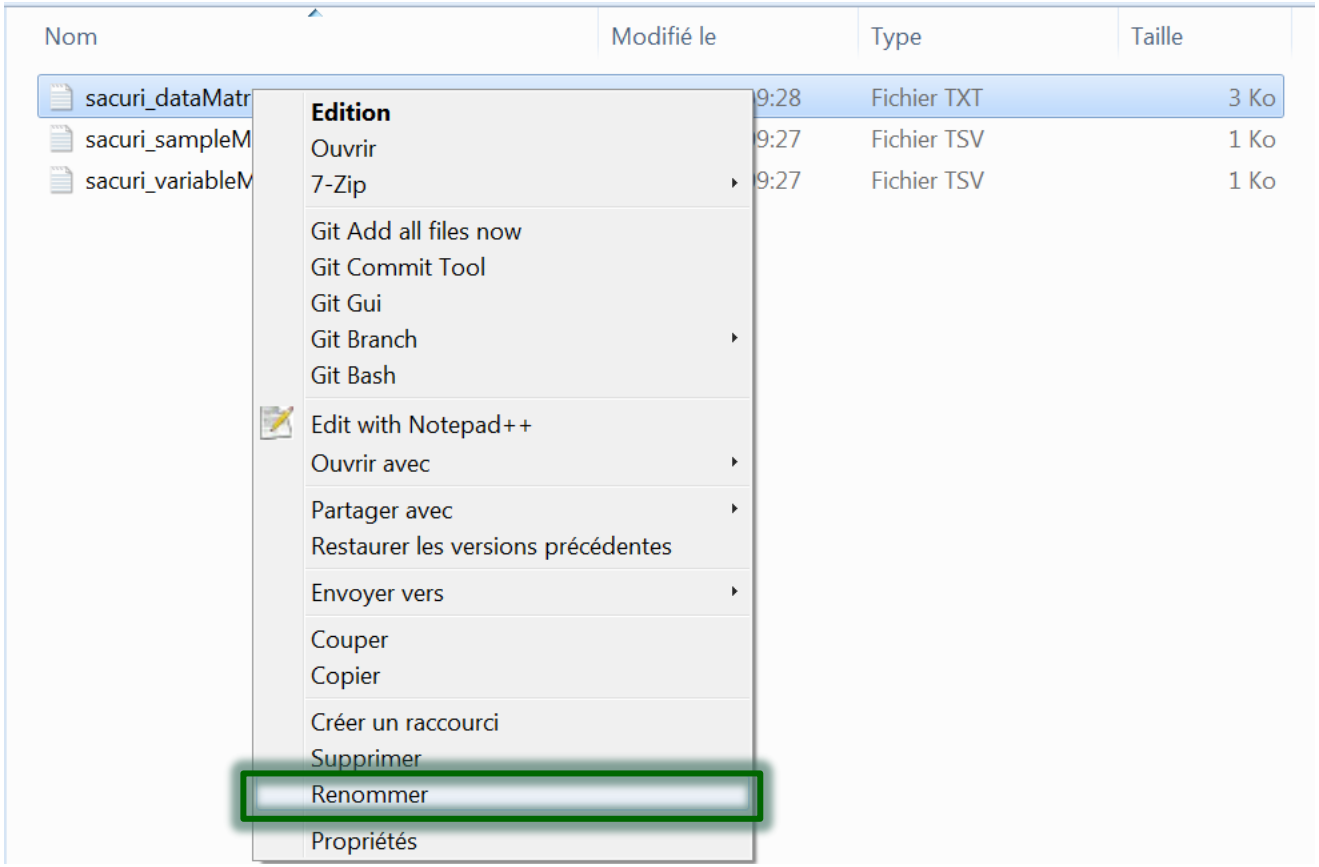
The input file format is TSV (tabulation separated values) or TABULAR. It is simple text file that can be prepared by using Excel and saved using the tabulated type format.



	1	2	3	4	5	6
1	dataMatrix	HU_neg_017	HU_neg_028	HU_neg_034	HU_neg_051	HU_neg_060
2	M97T61	17153667.17	10216240.88	16029523.86	14468044.45	8960121.78
3	M99T61	795428.1989	400570.6324	831219.0107	671471.606	334852.3461
4	M135T54	7057880.716	11926973.53	9514452.963	6990900.537	15613480.95
5	M136T54	221979.3107	223891.8602	311006.7896	223432.1903	389823.1439
6	M187T53	1087039.01	1011556.065	1315878.063	735998.5054	1682321.152
7	M189T53	306692.9525	271821.1866	356286.1817	205140.4975	436824.1702
8	M256T55	244723.3682	512382.4459	243893.1824	142140.6622	504325.7674
9	M111T64	812036.4987	3344894.025	1656818.089	481038.898	3026438.572
10	M99T318	762132.5736	2403252.338	795167.3119	412332.5154	2179804.785
11	M145T349	246841.7543	1105845.93	607635.7676	294122.4186	1483692.619
12	M405T464	61417.44523	157550.8765	26187.73666	71033.88904	157524.3854
13	M132T56	509220.2449	699447.5149	515630.0182	389402.7195	752634.908
14	M263T349	16892538.48	58849168.52	36829903.59	15886345.78	67075152.76
15	M289T65	1066673.164	2623407.553	1869960.276	420732.7062	1555054.973
16	M264T349	2105497.552	8012705.845	4982247.747	1926440.325	9254926.317
17	M167T64	4851369.143	1279443.724	3089285.412	3978500.899	1919598.755
18	M85T64	176339.1268	947481.7469	562753.9498	181842.1702	1011515.795
19	M289T51	119335.6511	50072.83283	125567.3167	121601.9527	55611.71589
20	M195T66	4743608.452	3848592.913	4644260.801	3191647.471	2006400.214

PREPARING YOUR FILES (2/3)

You can then rename your file with the .tsv extension (instead of .txt) by right-clicking on the file (and ignoring the warning):

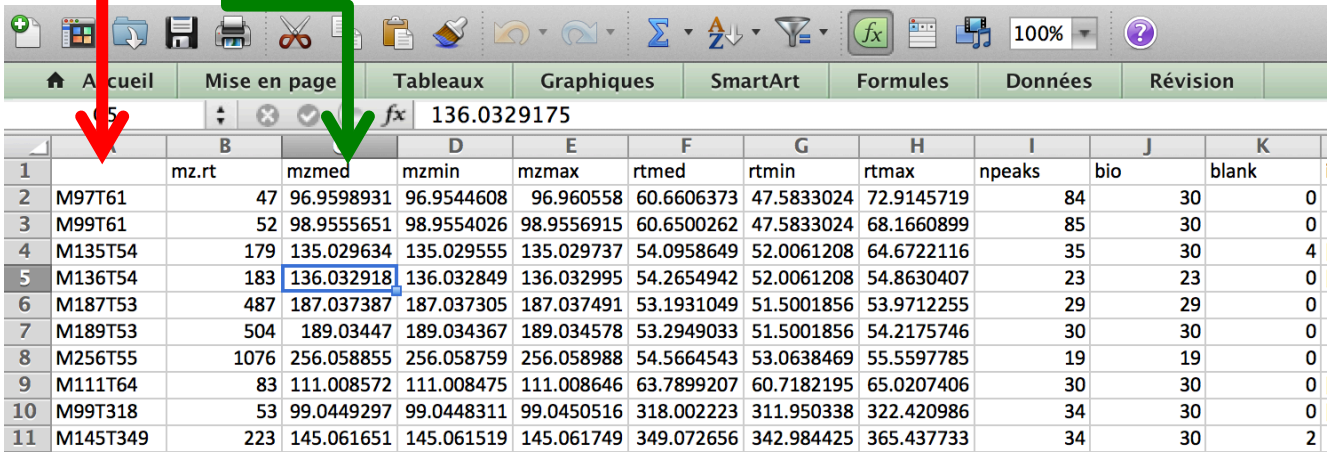


PREPARING YOUR INPUT FILE (3/3)

Such « .tsv » files (i.e. tabular separated; e.g. variableMetadata.tsv) can be handled correctly both by Excel and Galaxy.

The input file structure must contain:

- column with ID of the ion - example : M(mz)T(rt)
- column with masses of ions
- Decimal separator must be "."



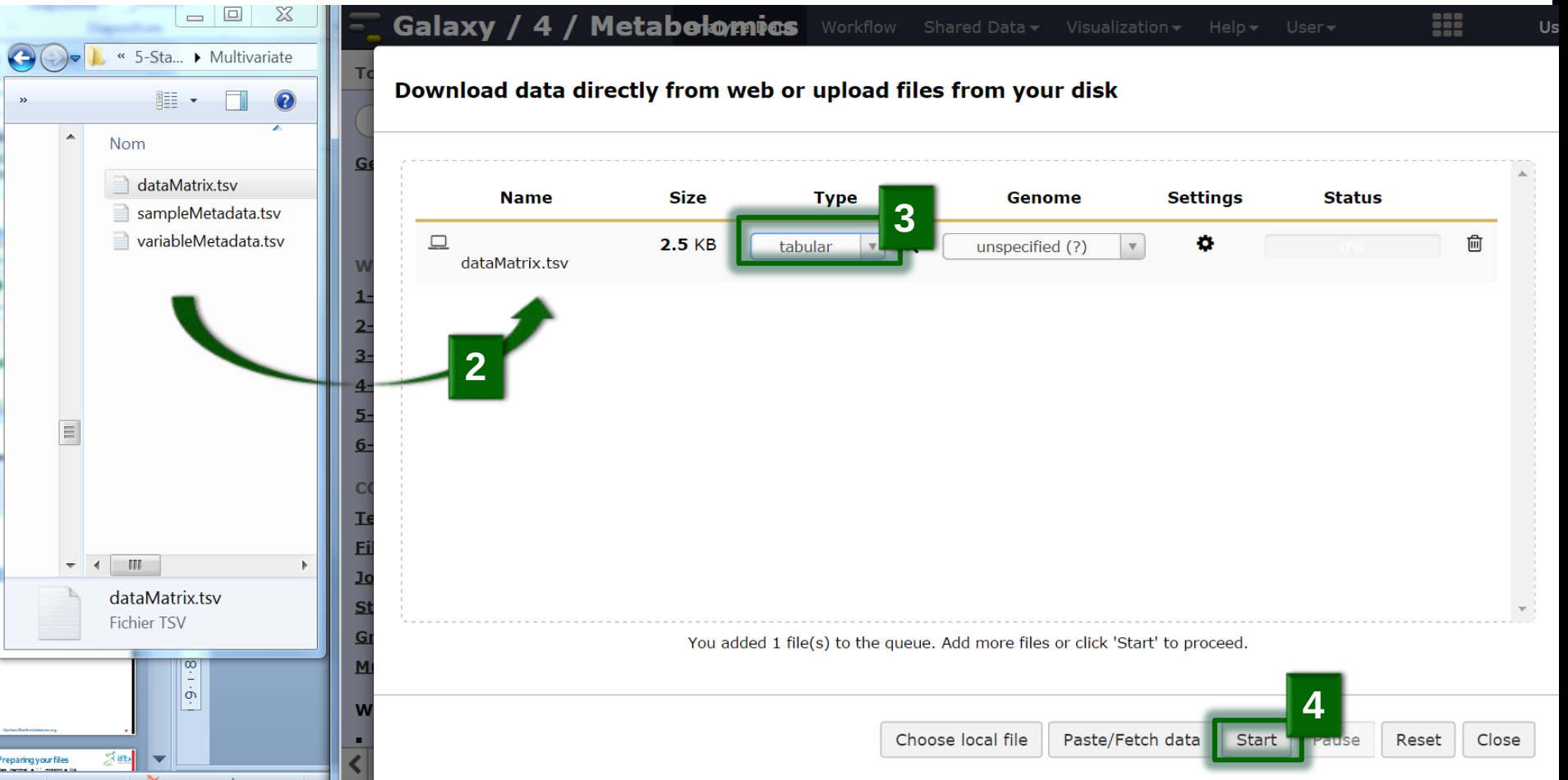
		mz.rt	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax	npeaks	bio	blank
2	M97T61	47	96.9598931	96.9544608	96.960558	60.6606373	47.5833024	72.9145719	84	30	0
3	M99T61	52	98.9555651	98.9554026	98.9556915	60.6500262	47.5833024	68.1660899	85	30	0
4	M135T54	179	135.029634	135.029555	135.029737	54.0958649	52.0061208	64.6722116	35	30	4
5	M136T54	183	136.032918	136.032849	136.032995	54.2654942	52.0061208	54.8630407	23	23	0
6	M187T53	487	187.037387	187.037305	187.037491	53.1931049	51.5001856	53.9712255	29	29	0
7	M189T53	504	189.03447	189.034367	189.034578	53.2949033	51.5001856	54.2175746	30	30	0
8	M256T55	1076	256.058855	256.058759	256.058988	54.5664543	53.0638469	55.5597785	19	19	0
9	M111T64	83	111.008572	111.008475	111.008646	63.7899207	60.7182195	65.0207406	30	30	0
10	M99T318	53	99.0449297	99.0448311	99.0450516	318.002223	311.950338	322.420986	34	30	0
11	M145T349	223	145.061651	145.061519	145.061749	349.072656	342.984425	365.437733	34	30	2

LOADING YOUR FILES INTO GALAXY (1/2)

Upload your file (variableMetadata.tsv)

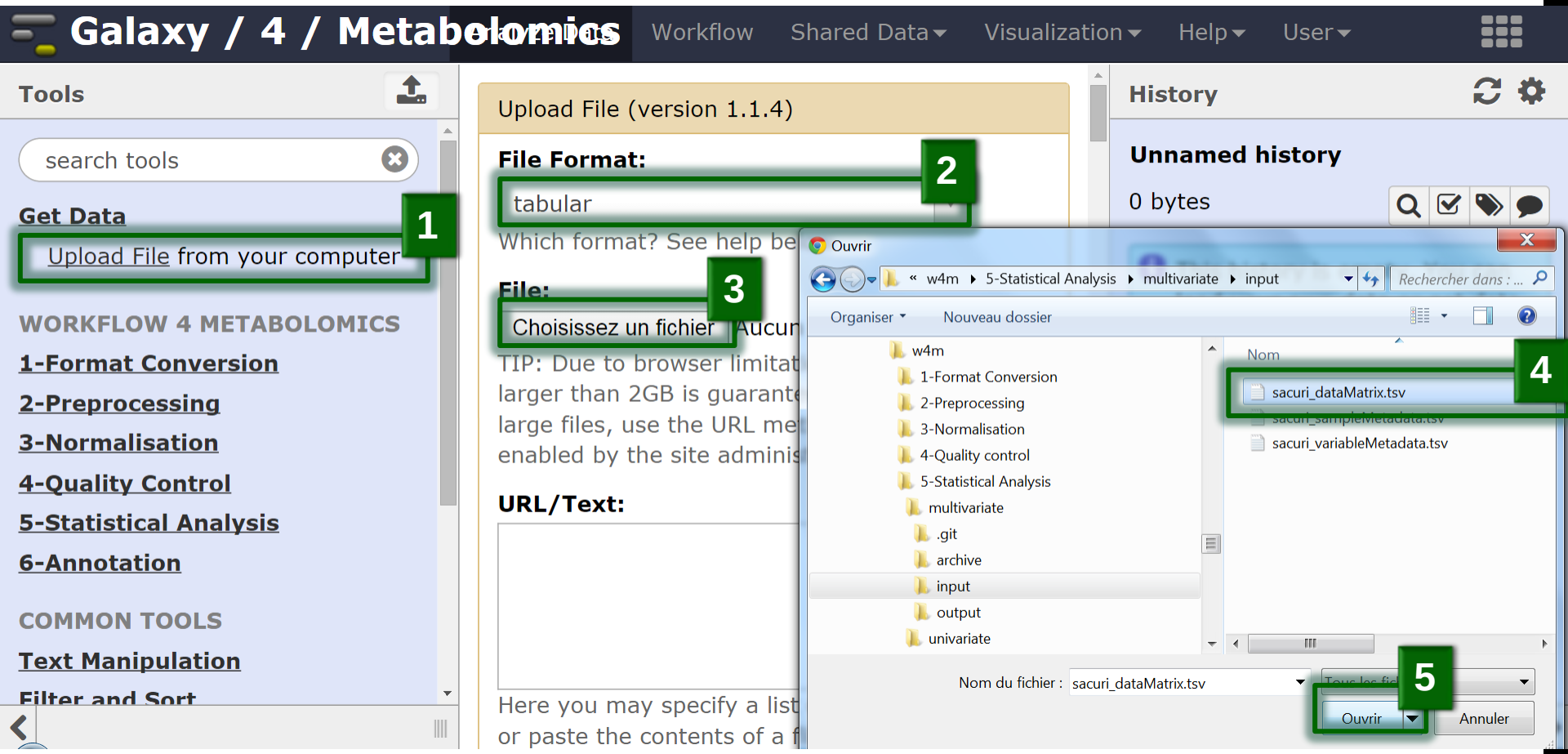
- either by using the icon 

and « drag & drop » the file:

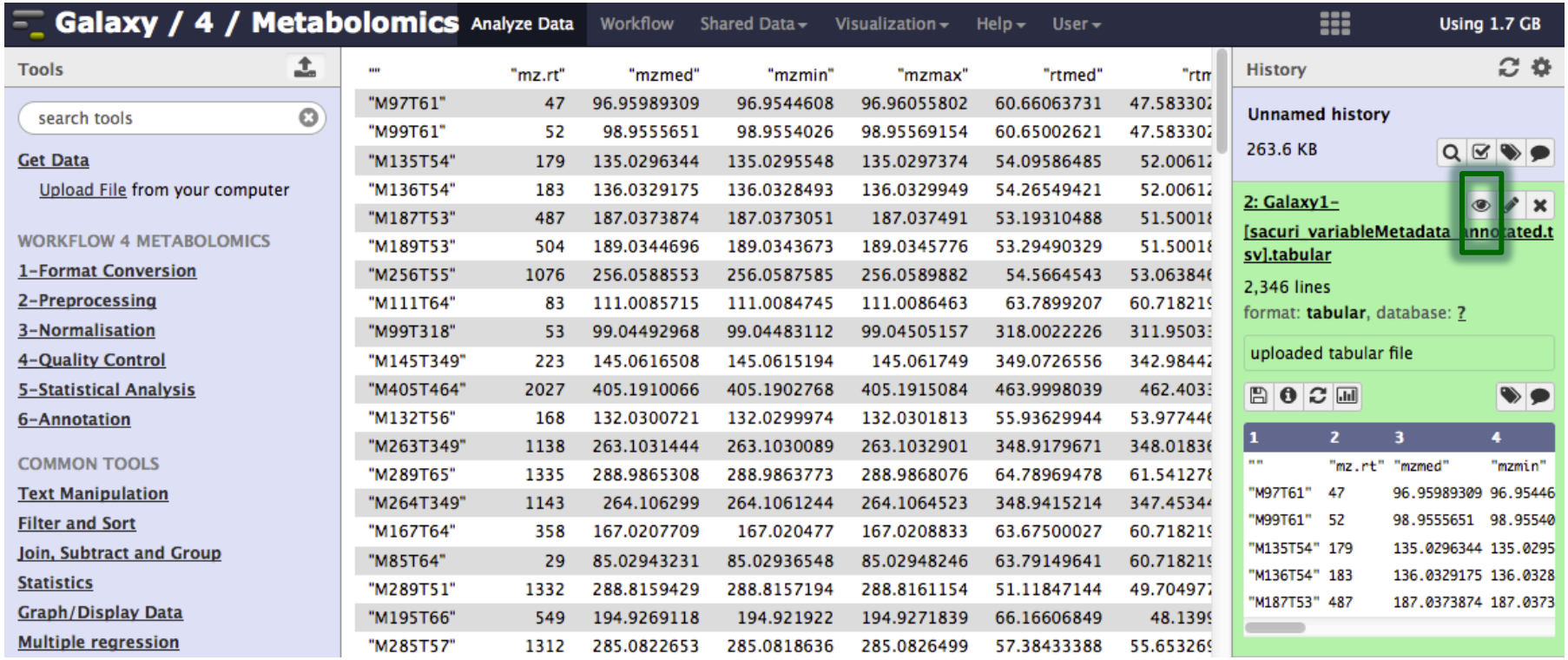


LOADING YOUR FILES INTO GALAXY (2/2)

- or with the Get Data / Upload File



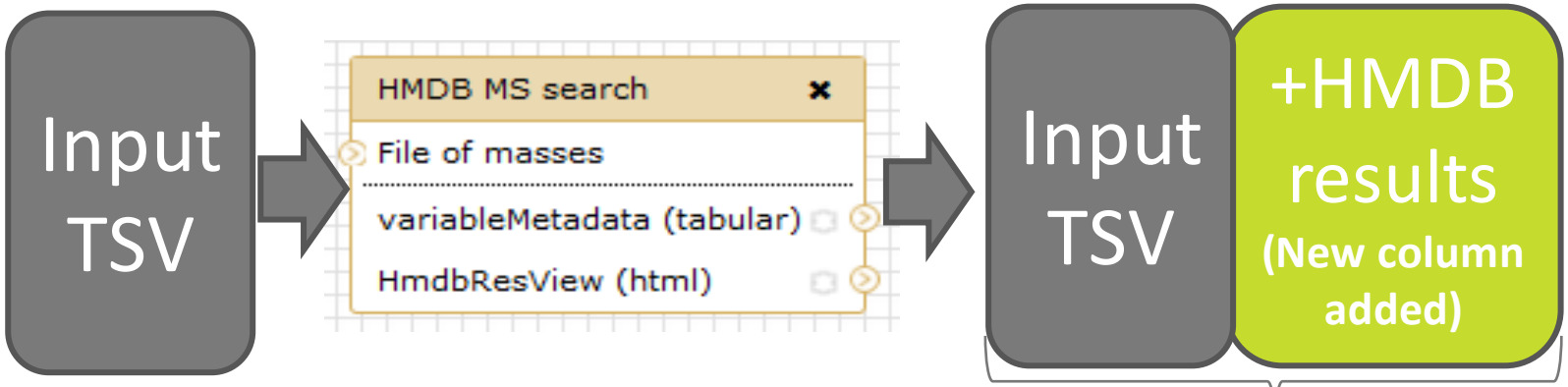
CHECK THAT YOUR DATA HAVE BEEN UPLOADED CORRECTLY



USE « LCMS ANNOTATION » MODULES

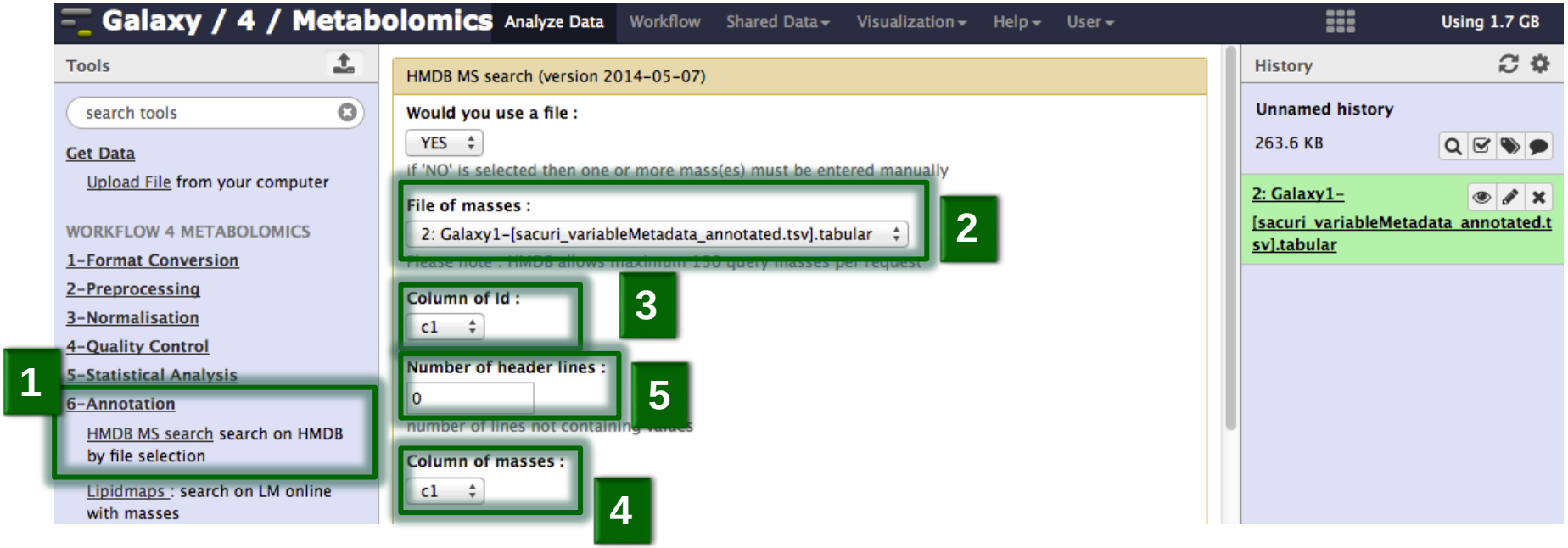
Each module proposes the same input and output formats, and the same type of parameters :

- **The common input and output formats** allow you to run multiple annotation tools sequentially or in parallel.
- **Commons outputs :**
  - HTML view: includes hyperlinks; multiple annotation results for the same variable are split into consecutive rows; next releases will provide more interactivity with the results.
  - .tsv format :



USE « LCMS ANNOTATION » MODULES

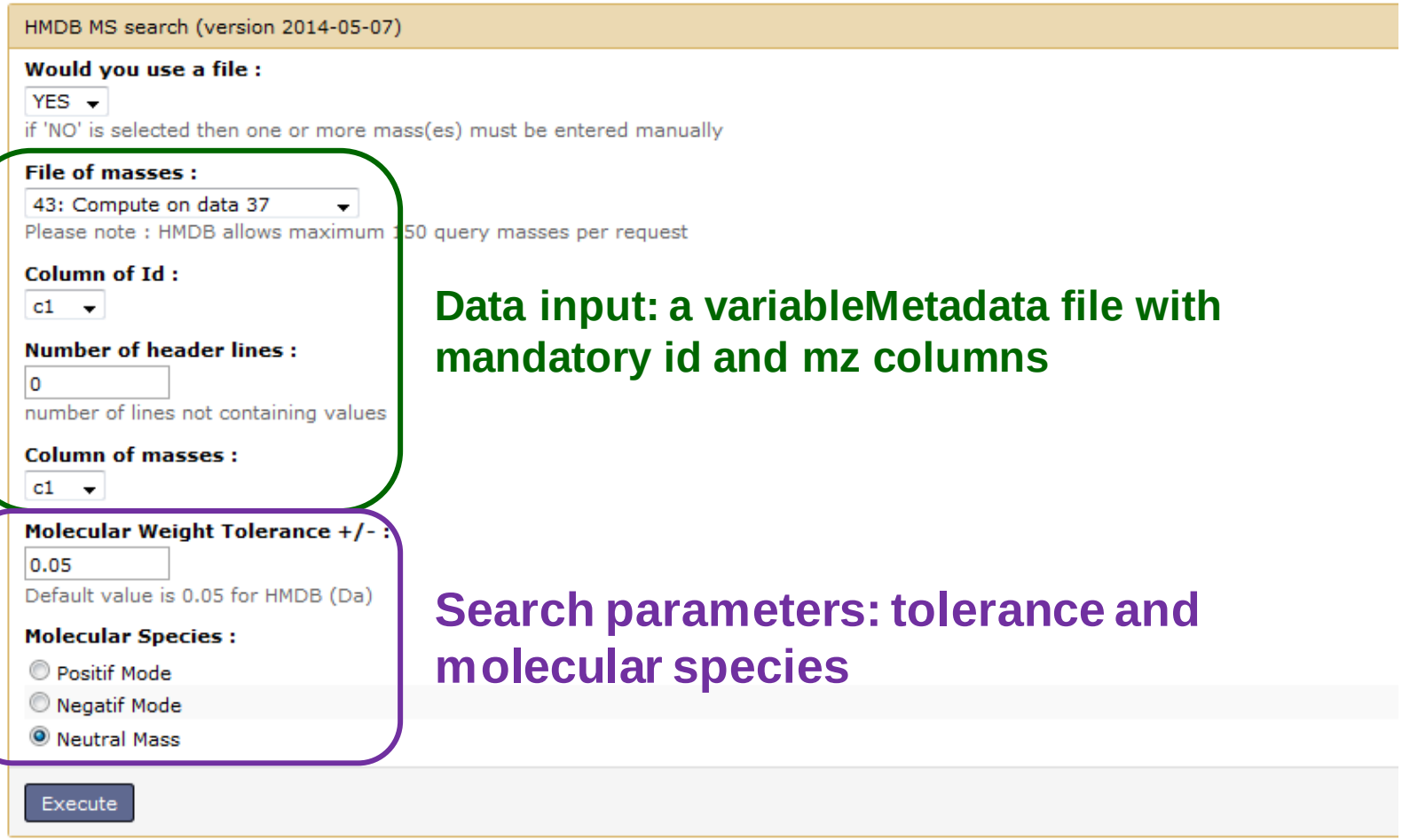
Open the module of your choice and select your input file.
Select position of your id and mass columns.



Indicate if your file contains header lines.

You are now ready to configure your annotation tool!

Example of HMDB input and configuration :



Data input: a variableMetadata file with mandatory id and mz columns

Search parameters: tolerance and molecular species

OUTPUT FORMATS

Annotation tool results – HTML view

Metadata of your query

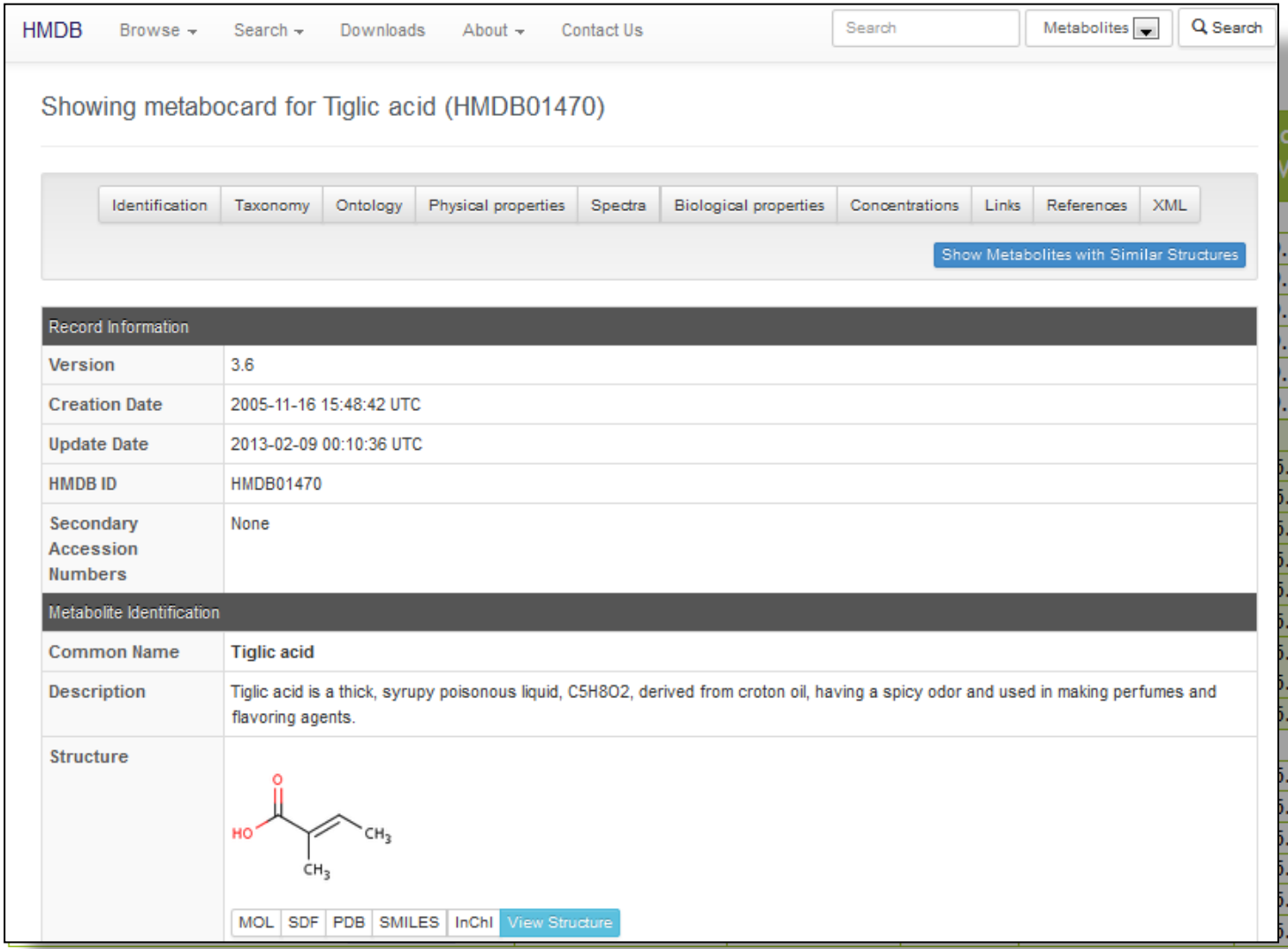
Results of HMDB queries - Search params : Molecular specie = negative / delta = 0.001

Prev 1 2 3 4 Next

ID from input	Mass (m/z)	Compound_ID	Formula	Compound MW (Da)	Adduct	Adduct_Type	Adduct MW (Da)	Delta
M99T318	99.04492968							
		HMDB41792	C4H6	54.04695	M+FA-H	-	99.045151	0.00022132
		HMDB01470	C5H8O2	100.05243	M-H	-	99.045154	0.00022432
		HMDB32351	C5H8O2	100.05243	M-H	-	99.045154	0.00022432
		HMDB01862	C5H8O2	100.05243	M-H	-	99.045154	0.00022432
		HMDB32459	C5H8O2	100.05243	M-H	-	99.045154	0.00022432
		HMDB30989	C10H16O4	200.104859	M-2H	-	99.045154	0.00022432
M145T349	145.0616508							
		HMDB33952	C4H8N2O	100.063663	M+FA-H	-	145.061864	0.0002132
		HMDB31642	C4H8N2O	100.063663	M+FA-H	-	145.061864	0.0002132
		HMDB02031	C5H10N2O3	146.069142	M-H	-	145.061866	0.0002152
		HMDB00641	C5H10N2O3	146.069142	M-H	-	145.061866	0.0002152
		HMDB06899	C5H10N2O3	146.069142	M-H	-	145.061866	0.0002152
		HMDB28687	C5H10N2O3	146.069142	M-H	-	145.061866	0.0002152
		HMDB03423	C5H10N2O3	146.069142	M-H	-	145.061866	0.0002152
		HMDB40613	C26H30O6	438.204239	M-3H	-	145.060804	0.0008468
		HMDB36597	C26H30O6	438.204239	M-3H	-	145.060804	0.0008468
M405T464	405.1910066							
		HMDB39013	C22H32O8	424.209718	M-H20-H	-	405.191328	0.0003214
		HMDB34492	C22H32O8	424.209718	M-H20-H	-	405.191328	0.0003214
		HMDB40795	C20H26O5	346.178024	M+Hac-H	-	405.191875	0.0008684
		HMDB12613	C21H28O5	360.193674	M+FA-H	-	405.191875	0.0008684
		HMDB41795	C20H26O5	346.178024	M+Hac-H	-	405.191875	0.0008684
		HMDB36901	C20H26O5	346.178024	M+Hac-H	-	405.191875	0.0008684

OUTPUT FORMATS

Annotation tool results – HTML links



OUTPUT FORMATS

.tsv output format: multiple annotations for the same variable are concatenated within the same cell using "::" as separator (Metlin format like)

	mz.rt	mzmed	rtmed	npeaks	bio	blank	isotopes	adduct	pcgroup	hmdb
M97T61	47	96.95989309	60.66063731	84	30	0		NA	57	"No_result_found_on HMDB"
M99T61	52	98.9555651	60.65002621	85	30	0		NA	57	"No_result_found_on HMDB"
M135T54	179	135.0296344	54.09586485	35	30	4	[13][M]+	NA	81	0.0002606::(C3H6O3)::HMDB01882|0.0002606::(C3H6O3
M136T54	183	136.0329175	54.26549421	23	23	0	[13][M+1]+	NA	81	0.0001435::(C10H14N2O7)::HMDB28818
M187T53	487	187.0373874	53.19310488	29	29	0		NA	138	0.0002726::(C9H10O3)::HMDB32030|0.0002726::(C9H10
M189T53	504	189.0344696	53.29490329	30	30	0		NA	413	"No_result_found_on HMDB"
M256T55	1076	256.0588553	54.5664543	19	19	0		NA	81	0.0002297::(C10H9N)::HMDB33115|0.0002297::(C10H9N
M111T64	83	111.0085715	63.7899207	30	30	0	[6][M]+	NA	4	0.0001965::(C5H4O3)::HMDB00617|0.0001965::(C5H4O3
M99T318	53	99.04492968	318.0022226	34	30	0	[3][M]+	NA	164	0.00022132::(C4H6)::HMDB41792|0.00022432::(C5H8O2
M145T349	223	145.0616508	349.0726556	34	30	2		NA	5	0.0002132::(C4H8N2O)::HMDB33952|0.0002132::(C4H8N
M405T464	2027	405.1910066	463.9998039	29	29	0	[219][M]+	NA	175	0.0003214::(C22H32O8)::HMDB39013|0.0003214::(C22H
M132T56	168	132.0300721	55.93629944	30	30	0		NA	109	0.0001569::(C3H5NO2)::HMDB39426|0.0001569::(C3H5N
M263T349	1138	263.1031444	348.9179671	35	30	5	[121][M]+	NA	5	0.0000376::(C13H18N2O5)::HMDB61146|0.0000376::(C1
M289T65	1335	288.9865308	64.78969478	30	30	0		NA	167	"No_result_found_on HMDB"
M264T349	1143	264.106299	348.9415214	34	30	4	[121][M+1]+	NA	5	0.000072::(C13H17NS)::HMDB40032

"No_result_found_on HMDB"

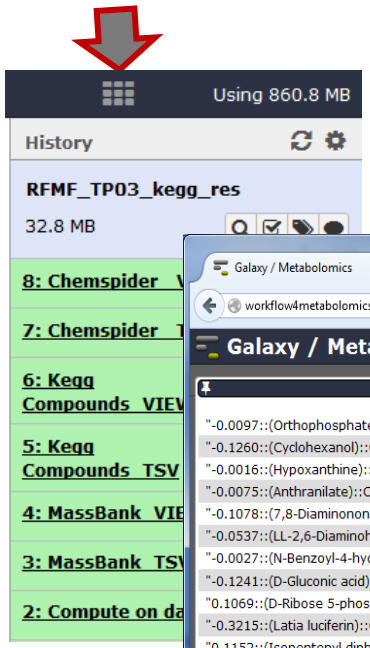
0.0002606::(C3H6O3)::HMDB01882|0.0002606::(C3H6O3

« METLIN FORMAT LIKE»

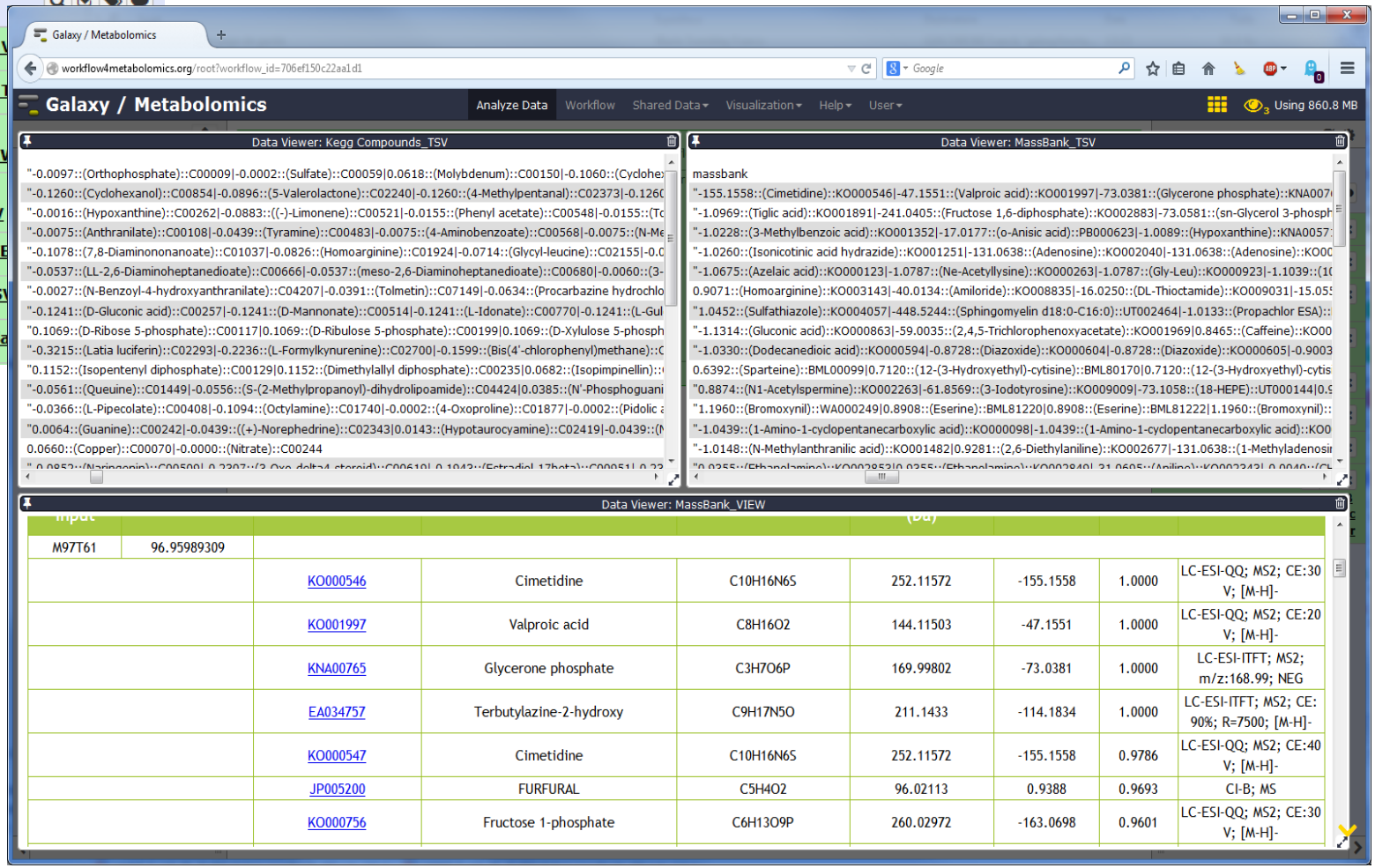
0.003604744502::(Clitocine)::HMDB33718
0.003605255498::(Citrusinine II)::HMDB30373

DELTA **COMPOUND_NAME** **HMDB_ID**

« COMPARE » ANNOTATION



You can use the multi-view tool of W4M
and visualize your annotation files simultaneously.



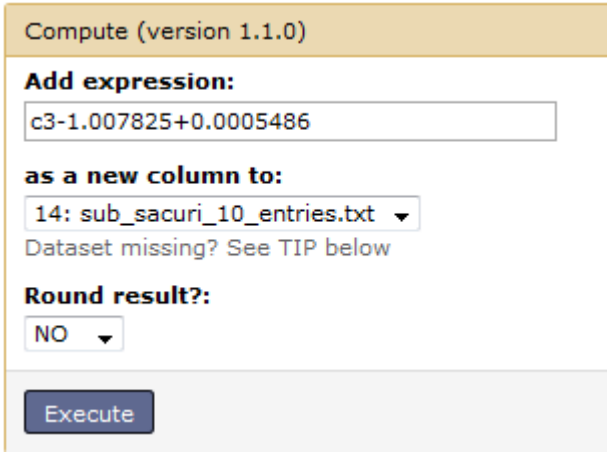
M97T61	96.95989309							
		KO000546	Cimetidine	C10H16N6S	252.11572	-155.1558	1.0000	LC-ESI-QQ; MS2; CE:30 V; [M-H]-
		KO001997	Valproic acid	C8H16O2	144.11503	-47.1551	1.0000	LC-ESI-QQ; MS2; CE:20 V; [M-H]-
		KNA00765	Glycerone phosphate	C3H7O6P	169.99802	-73.0381	1.0000	LC-ESI-ITFT; MS2; m/z:168.99; NEG
		EA034757	Terbutylazine-2-hydroxy	C9H17N5O	211.1433	-114.1834	1.0000	LC-ESI-ITFT; MS2; CE: 90%; R=7500; [M-H]-
		KO000547	Cimetidine	C10H16N6S	252.11572	-155.1558	0.9786	LC-ESI-QQ; MS2; CE:40 V; [M-H]-
		JP005200	FURFURAL	C5H4O2	96.02113	0.9388	0.9693	CI-B; MS
		KO000756	Fructose 1-phosphate	C6H13O9P	260.02972	-163.0698	0.9601	LC-ESI-QQ; MS2; CE:30 V; [M-H]-

FORMAT YOUR MZ (IF NECESSARY)

positive / negative ionization modes:

- Some tools are not mass spectrometry oriented (ex: KEGG)
- It is therefore necessary to format your mz values with COMMON TOOLS / Compute prior to using them

  - positive mode: C3 - 1.007825 (Proton) + 0.0005486 (electron)
  - negative mode: C3 + 1.007825 (Proton) - 0.0005486 (electron)

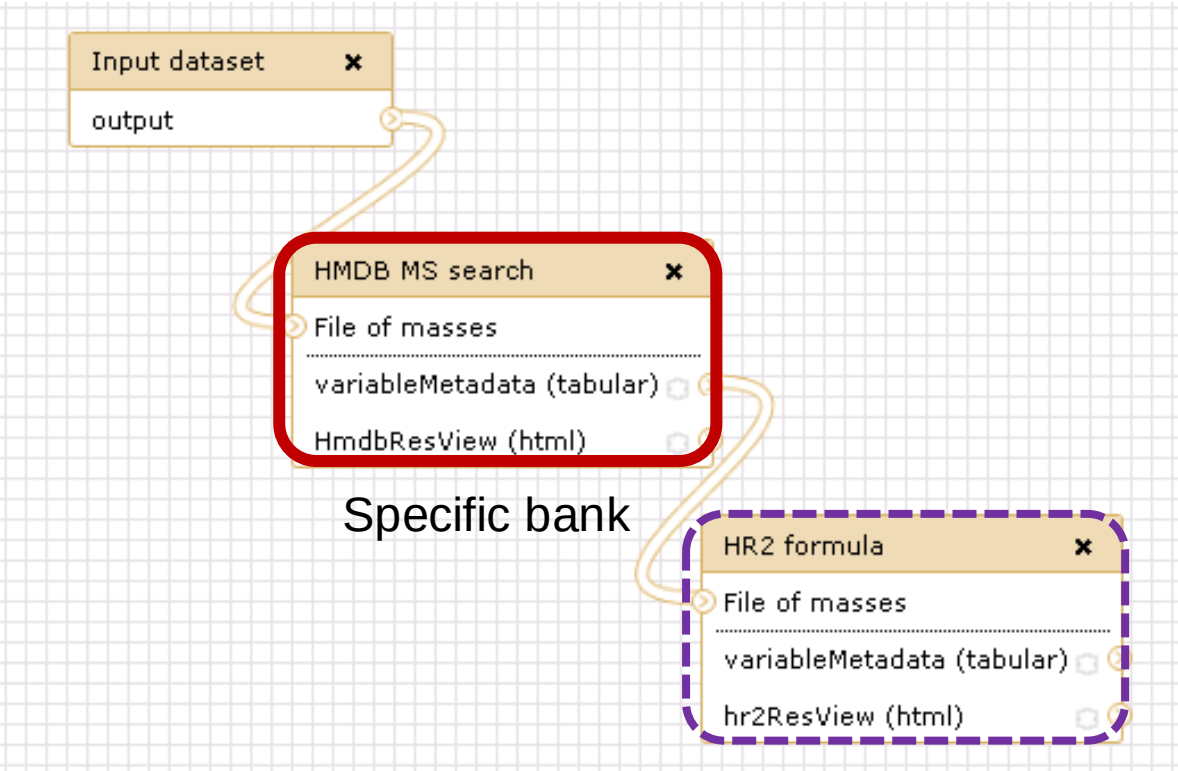


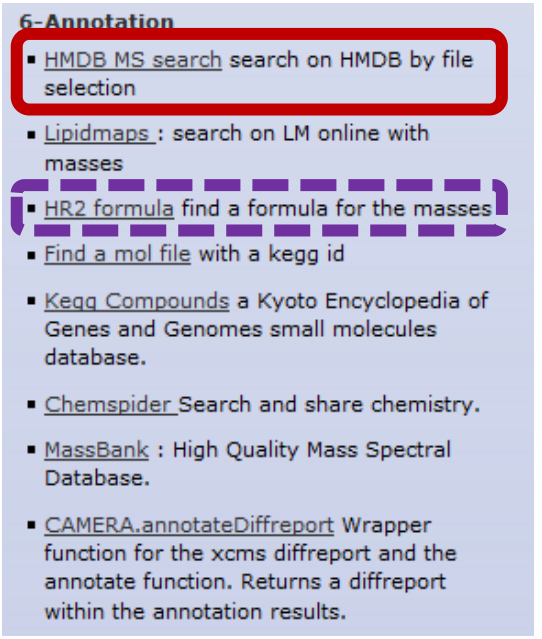
Mzmed [M+H]	M
132.0300721	131.0227957
263.1031444	262.095868
288.9865308	287.9792544
264.106299	263.0990226

ANNOTATION WORKFLOW BUILDING

Example of a sequential « Chemical » annotation workflow:

Chain a specific bank query and a formula elucidator



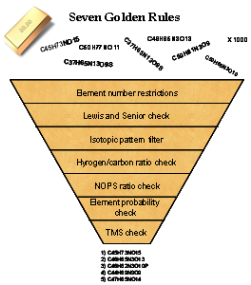


« CHEMICAL » ANNOTATION WORKFLOW FEATURES

Human Metabolome DataBase:

- Bank containing small molecule metabolites found in the human body.
- V3.6 with 41,815 entries + 5,688 proteins.

HR2:



- Molecular formula generator: HiRes ("High Resolution") MS by Joerg Hau.
- Generate random chemical formula and filter them with “the seven golden rules”

ANNOTATION WORKFLOW BUILDING

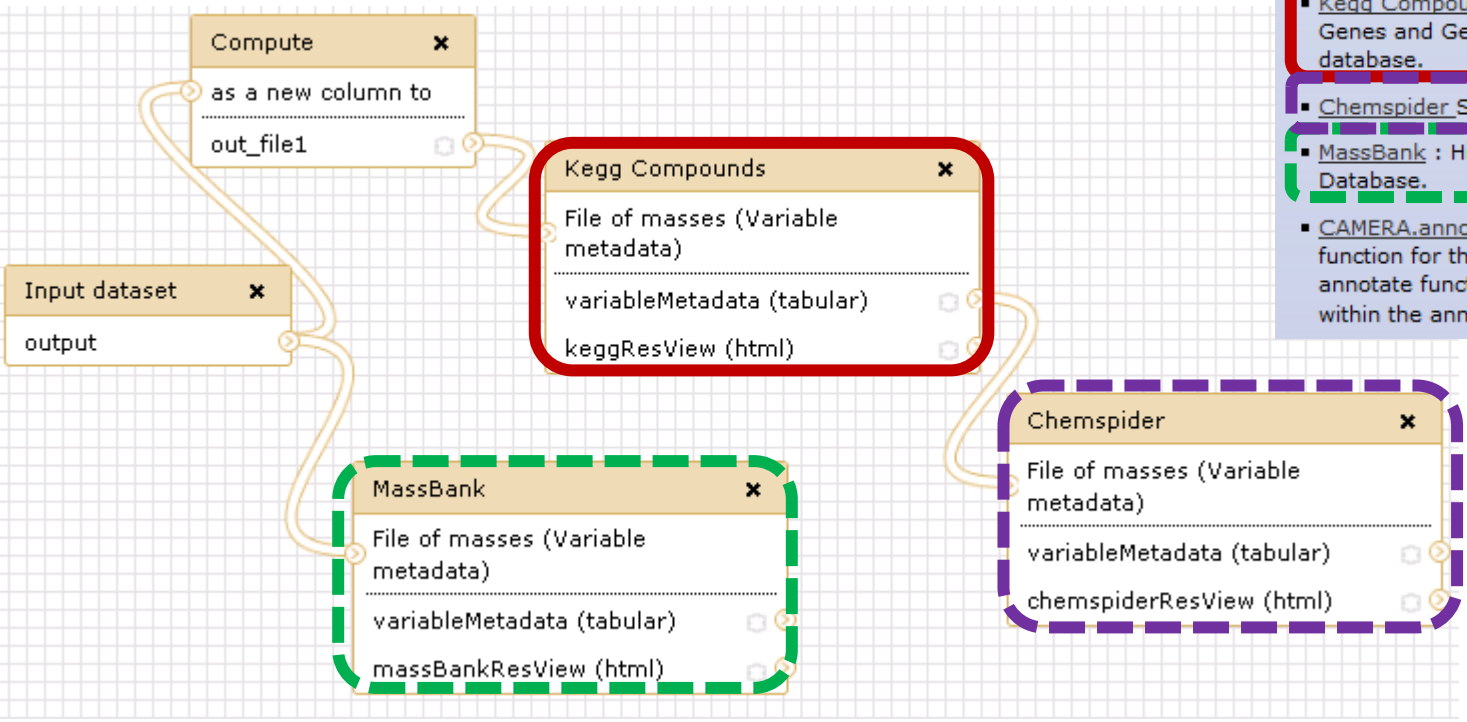
Example of a parallel annotation workflow:

Combine chemical compound libraries

and a spectral repository.

6-Annotation

- HMDB MS search search on HMDB by file selection
- Lipidmaps : search on LM online with masses
- HR2 formula find a formula for the masses
- Find a mol file with a kegg id
- Kegg Compounds a Kyoto Encyclopedia of Genes and Genomes small molecules database.
- Chemspider Search and share chemistry.
- MassBank : High Quality Mass Spectral Database.
- CAMERA.annotateDiffreport Wrapper function for the xcms diffreport and the annotate function. Returns a diffreport within the annotation results.



« CHEMICAL » ANNOTATIONS

KEGG Compounds:

- Sub-bank of KEGG containing small molecules, biopolymers, and other chemical substances, relevant to biological systems. V2014-05 with 17,254 entries.

MassBank:

- Public repository of mass spectra of small chemical compounds for life sciences.
- V2014-03 with 40,889 spectra for 15,775 compounds

ChemSpider

- Meta engine providing text and structure search to over 30 millions structures from ~500 data sources.

REFERENCES

- Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, et al., HMDB 3.0—The Human Metabolome Database in 2013. Nucleic Acids Res. 2013.
- Silva, R. R. and Jourdan, F. and Salvanha, D. M. and Letisse, F. and Jamin, E. L. and Guidetti-Gonzalez, S. and Labate, C. A. and Vencio, R. Z. N. (2014). ProbMetab: an R package for Bayesian probabilistic annotation of LC-MS-based metabolomics. In Bioinformatics, 30 (9), pp. 1336–1337.
- Sud M., Fahy E., Cotter D., Brown A., Dennis E., Glass C., Murphy R., Raetz C., Russell D., and Subramaniam S. LMSD: LIPID MAPS structure database. Nucleic Acids Research 35, D527-32 (2006).
- Tobias Kind and Oliver Fiehn. (2007). "Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry." BMC Bioinformatics p8:105
- Kanehisa, M., S. Goto, et al. (2012). "KEGG for integration and interpretation of large-scale molecular data sets." Nucleic Acids Research 40(D1): D109-D114
- Pence, H.E. and A.J. Williams, (2010). "ChemSpider: An Online Chemical Information Resource." J.Chem. Educ., 87(11): p. 1123-1124.
- H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa. Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Yokota Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka, (2010). "MassBank: A public repository for sharing mass spectral data for life sciences." J. Mass Spectrom., 45, 703-714.

END !