

## MS-DIAL tutorial:

### Universal program for untargeted metabolomics

14/04/2020

Justine Chervin and Guillaume Marti

[justine.chervin@lrsv.ups-tlse.fr](mailto:justine.chervin@lrsv.ups-tlse.fr)

[guillaume.marti@univ-tlse3.fr](mailto:guillaume.marti@univ-tlse3.fr)

<https://www.lrsv.ups-tlse.fr/metatoul-en/>

## 1. Prerequisite:

### Software installation:

**MS-DIAL** last version: <http://prime.psc.riken.jp/compms/msdial/main.html>

**Abf (Analysis Base File) Converter:** <https://www.reifycs.com/AbfConverter/>

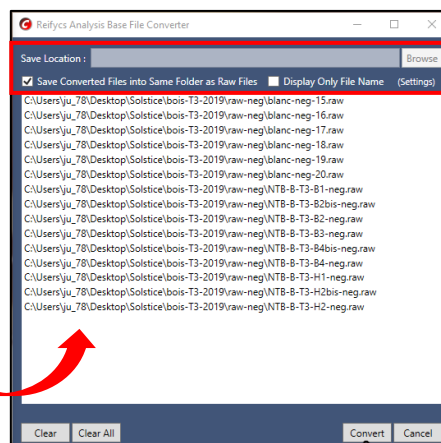
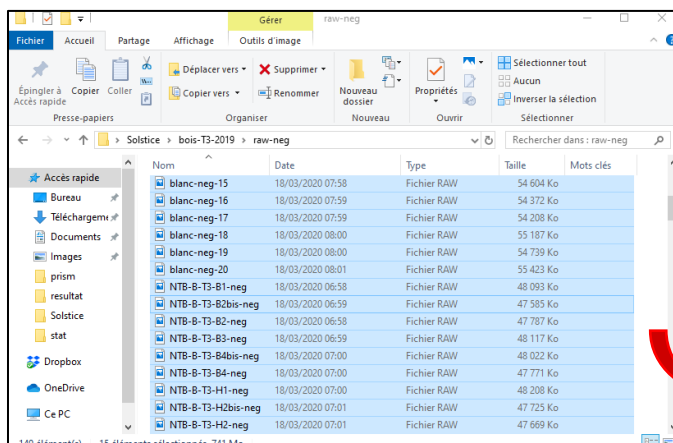
**MSFileReader** for Thermo data: see the following link which explains how to process <http://fields.scripps.edu/rawconv/>

## 1. File conversion:

- Start the application AnalysisBaseFileConverter
- Drag and drop** .raw files into the application
- Choose the folder of arrivals (cf. red rectangle)

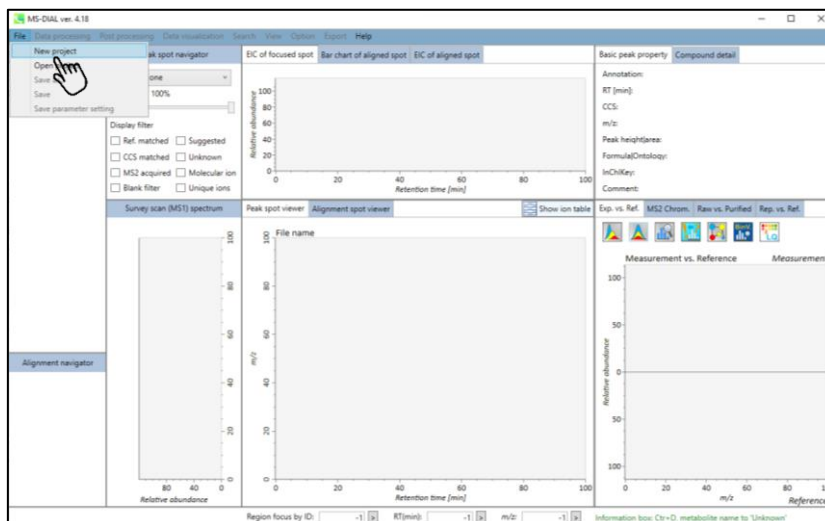
By default, the .abf files will be created in the same folder as raw data since « **Save Converted Files into the Same Folder as Raw Files** » is selected. If you want to generate them into another folder, uncheck this and select the new folder by clicking « **Browse** » on the top right.

- Click on « **Convert** »

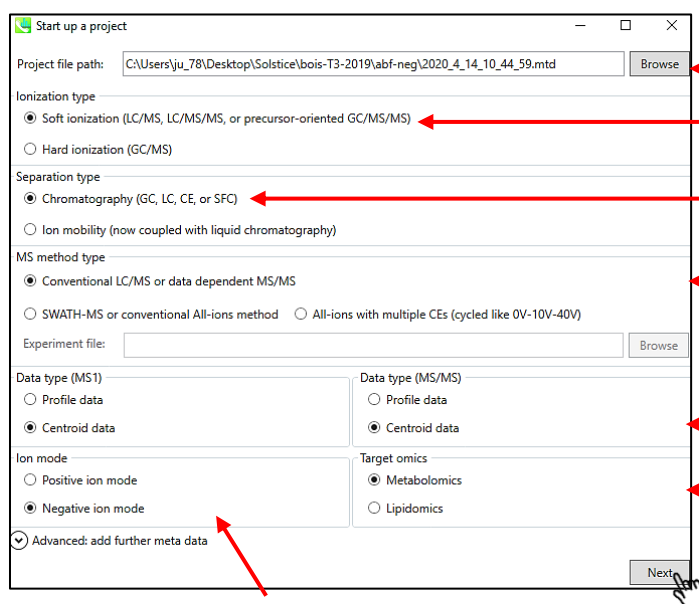


## 2. Create a project on MS-DIAL and Import data

- a. Open the application MS-DIAL and on the « **File** » menu select « **New project** »



The following window appears to define the type of experiments you are about to process:



Select the folder containing the .abf files

Select « Soft ionisation » for ESI experiments  
For EI GC/MS select “Hard ionization”

Select the type of chromatography used

Select « Conventional LC/MS » if you want to process LC/MS or LC/MS/MS data in DDA mode  
Select SWATCH-MS\* if you want to process LC/MS/MS data in DIA mode

Select Profile or Centroid data

Select « Metabolomics » to use your own MSP file database otherwise « Lipidomics » to use the internal lipid database of MS-DIAL

Select the ion mode (you have to process pos and neg separately)

\* by selecting SWATCH-MS, you have to indicate a “dictionary file” as tab delimited text file with MS1 scan range and precursor window as follows:

<http://prime.psc.riken.jp/compms/msdial/main.html#Templates>

## b. Import your data

The following window appears and click on “**Browse**” to import the .abf files you want to process. During data importation, it is important to note the type (Blank, QC or Sample) and class of every sample in **Class ID column** (blank, sample class, QC). If both ionization modes have been acquired, be careful to have the **same number of samples** between pos and neg modes and in the **same order**.

New project window

Analysis file paths

File path	File name	Type	Class ID	Batch	Analytical order	Inject. volume (μL)	Included
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	blanc-ext-neg1	Sample	blancext	1	1	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	blanc-ext-neg2	Sample	blancext	1	2	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	blanc-neg-15	Blank	Blank	1	3	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	blanc-neg-16	Blank	Blank	1	4	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	blanc-neg-17	Blank	Blank	1	5	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-B1-neg	Sample	NTB-B-B	1	6	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-B2-neg	Sample	NTB-B-B	1	7	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-B3-neg	Sample	NTB-B-B	1	8	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-B4-neg	Sample	NTB-B-B	1	9	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-H1-neg	Sample	NTB-B-H	1	10	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-H2-neg	Sample	NTB-B-H	1	11	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-H3-neg	Sample	NTB-B-H	1	12	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-H4-neg	Sample	NTB-B-H	1	13	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-I1-neg	Sample	NTB-B-I	1	14	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-I2-neg	Sample	NTB-B-I	1	15	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-I4-neg	Sample	NTB-B-I	1	16	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	QC-T3-neg1	QC	QC	1	17	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	QC-T3-neg2	QC	QC	1	18	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	QC-T3-neg3	QC	QC	1	19	1	<input checked="" type="checkbox"/>
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	QC-T3-neg4	QC	QC	1	20	1	<input checked="" type="checkbox"/>

## 3. Data analysis and parameters setting

### a. Data collection parameters

Indicate the MS1 and MS2 tolerance  
Depend of your instrument

Indicate the range of RT

Indicate the range of MS1 mass

Indicate the range of MS2 mass

For small molecules, the maximum charged number can stay at 2 but you can go up to 8 if necessary.

You can fix the count of threads used to process data (see resource monitor in administrative tools (control panel))

## b. Peak detection parameters

## Fix the peak height threshold\*

- By default, 0.1 is generally suitable

Linear weighted moving average is set by default with recommended smoothing level 1-3

Fix the peak width threshold (by default 5 is ok)

You can specify some unwanted peaks if necessary

**\* Critical value which greatly influence final results: 70% under the observed baseline is a good starting point.**

### c. MS2Dec parameters

The screenshot shows the 'Analysis parameter setting' dialog box with the 'MS2Dec' tab selected. The 'Deconvolution parameters' section includes a 'Sigma window value' set to 0.5 and an 'MS/MS abundance cut off' set to 10. The 'Advanced' section has 'Exclude after precursor ion' checked, 'Keep the isotopic ions until' set to 0.5 Da, and 'Keep the isotopic ions w/o MS2Dec' checked. At the bottom, there are 'Load', 'Together with Alignment' (checked), 'Finish', and 'Cancel' buttons.

By default, 0.5 is suitable. Under 0.3, you will recognize noise chromatographic peaks. Up to 0.7, you will reduce peak resolution

Set a cut off value to reduce the MS/MS noise

Recommended for metabolomics and lipidomics

Set at 0.5 by default.

Select this because the algorithm may erase the precursor's isotopic ions.

### d. Identification parameters

The screenshot shows the 'Analysis parameter setting' dialog box with the 'Identification' tab selected. The 'MSP file and MS/MS identification setting' section includes an 'MSP file' field with a path to a .msp file and a 'Select' button. Below are 'Retention time tolerance' (0.1 min), 'Accurate mass tolerance (MS1)' (0.01 Da), 'Accurate mass tolerance (MS2)' (0.05 Da), and 'Identification score cut off' (80 %). The 'Advanced' section has 'Use retention time for scoring' and 'Use retention time for filtering' both checked. The 'Text file and post identification (retention time and accurate mass based) setting' section includes a 'Text file' field and a 'Select' button. Below are 'Retention time tolerance' (0.1 min), 'Accurate mass tolerance' (0.01 Da), and 'Identification score cut off' (85 %). The 'Spectrum cut off and report option' section has 'Relative abundance cut off' (0 %) and 'Only report the top hit' (unchecked). At the bottom, there are 'Load', 'Together with Alignment' (checked), 'Finish', and 'Cancel' buttons.

You can select a msp format database for identification of level 1. It contains information such as accurate mass, retention time, MS/MS peaks and various identifiers.

Set the mass and retention time tolerances.

The identification score cut off should be greater than 70%.

You can select a text format database for retention time and accurate mass search (post identification)

The first two required columns should contain the "metabolite name" and the "accurate mass". You can add additional information such as the retention time in following columns. If so, set a suitable retention time tolerance, otherwise indicate "100".

The identification score cut off should be greater than 70%.

## Example of msp format database:

DB-IRSV-neg-9420 - Bloc-notes

Fichier Edition Format Affichage Aide

NAME: (Metasci) trans-Resveratrol

SCA NUMBER: 174

RETENTIONTIME: 11.77932

PRECURSORMZ: 227.0715

PRECURSOR TYPE: [M-H]<sup>-</sup>

IONMODE: Negative

SPECTRUMTYPE: Centroid

FORMULA: C14H12O3

INCHIKEY: LUKBXSANLPMWSZ-OMJ3BTEDSA-N

INCHI: OC1=CC=C(C=C1)C=CC2=CC(O)=CC(O)=C2

SMILES: OC1=CC=C(C=C1)C=CC2=CC(O)=CC(O)=C2

AUTHORS:

COLLISIONENERGY: 40

INSTRUMENT:

INSTRUMENTTYPE:

IONIZATION:

LICENSE:

COMMENT: 174

Num Peaks: 16

m/z	Intensity	Theoretical m/z	Mass diff (ppm)	Formula
117.03439	15932	117.034591	1.72	SMILES OC1=CC=C(C=C1)C=CC2=CC(O)=CC(O)=C2
119.05083	14844	119.050242	1.78	SMILES OC1=CC=C(C=C1)C=CC2=CC(O)=CC(O)=C2
138.96481	7961			
141.0718521		141.070425	0	Formula C11H9 <sup>+</sup>
143.0502	289851	143.04969	0	Formula C10H7O <sup>+</sup>
144.05347	26946			
145.02922	8625	145.028954	0	Formula C9H5O2 <sup>+</sup>
157.06575	28892	157.06534	0	Formula C11H9O <sup>+</sup>
159.08139	39051	159.08099	0	Formula C11H11O <sup>+</sup>
181.06555	19544	181.06534	0	Formula C13H9O <sup>+</sup>
182.07332	24964			
182.95444	9139			
183.08125	55758	183.08099	0	Formula C13H11O <sup>+</sup>
185.06058	196543	185.060255	0	Formula C12H9O2 <sup>+</sup>
186.06398	27923			
227.07161	780266	227.071367	0	SMILES OC1=CC=C(C=C1)C=CC2=CC(O)=CC(O)=C2

NAME: (Metasci) cis-Resveratrol

SCA NUMBER: 174

RETENTIONTIME: 12.491

PRECURSORMZ: 227.0715

PRECURSOR TYPE: [M-H]<sup>-</sup>

IONMODE: Negative

SPECTRUMTYPE: Centroid

## Example of text format database:

msdiaDB-txt - Bloc-notes

Fichier Edition Format Affichage Aide

name	MZ	RT
(2-AMINOETHYL)PHOSPHONATE	124.0168	1.146
10-HYDROXYDECANOATE	187.1336	12.42378
1-HYDROXY-2-NAPHTHOATE	187.0399	9.168233
2,3-DIHYDROXYBENZOATE	153.0192	6.998233
2',4'-DIHYDROXYACETOPHENONE	151.04	11.05613
2,4-DIHYDROXYPTERIDINE	163.0259	4.9144
2,5-DIHYDROXYBENZOATE	153.0193	7.91735
2-AMINOISOBUTYRATE	102.0558	1.300117
2-AMINOPHENOL	108.0452	2.778317
2-HYDROXY-4-(METHYLTHIO)BUTANOATE	149.0276	6.95675
2-HYDROXYBUTYRATE	103.0399	3.981767
2-HYDROXYPHENYLACETATE	151.0399	7.821167
2-METHYLGLUTARATE	145.0506	7.2758
3-(2-HYDROXYPHENYL)PROPANOATE	165.0553	10.48115
3,4-DIHYDROXYMANDELATE	183.0296	2.593883
3,4-DIHYDROXYBENZOATE	153.0191	6.994184
3,4-DIHYDROXYPHENYLACETATE	167.0347	3.969783
3,4-DIHYDROXYPHENYLGLYCOL	169.0504	3.22905
3,5-DIODO-L-TYROSINE	523.8864	11.18588
3,5-DIODO-L-TYROSINE	431.8596	8.64365
3-AMINO-4-HYDROXYBENZOATE	152.0352	2.763383
3-AMINO-5-HYDROXYBENZOATE	152.0352	3.083767
3-DEHYDROSHIKIMATE	171.0296	2.2125
3-HYDROXYBENZOATE	137.0244	8.760616
3-HYDROXYBUTANOATE	103.0399	8.238883
3-HYDROXYMETHYLGLUTARATE	161.0454	4.483967
3-METHOXY-4-HYDROXYMANDELATE	151.0399	8.7933
3-METHOXYTYROSINE	197.0453	5.3751

## e. Adduct ion setting

Analysis parameter setting

Data collection Peak detection MS2Dec Identification Adduct Alignment Mobility Isotope tracking

Adduct ion setting

User-defined adduct

Molecular species	Charge	Accurate mass [Da]	Included
[M-H] <sup>-</sup>	1	-1.00782503207	<input checked="" type="checkbox"/>
[M-H2O-H] <sup>-</sup>	1	-19.01838971207	<input checked="" type="checkbox"/>
[M+Na-2H] <sup>-</sup>	1	20.97411921676	<input type="checkbox"/>
[M+Cl] <sup>-</sup>	1	34.96885268	<input type="checkbox"/>
[M+K-2H] <sup>-</sup>	1	36.94805661586	<input type="checkbox"/>
[M+FA-H] <sup>-</sup>	1	44.99765396793	<input checked="" type="checkbox"/>
[M+Hac-H] <sup>-</sup>	1	59.01330396793	<input type="checkbox"/>
[M+C2H3N+Na-2H] <sup>-</sup>	1	62.00066831777	<input type="checkbox"/>
[M+Br] <sup>-</sup>	1	78.9183371	<input type="checkbox"/>
[M+TFA-H] <sup>-</sup>	1	112.98503896793	<input type="checkbox"/>
[M-C6H10O4-H] <sup>-</sup>	1	-147.06573383101	<input type="checkbox"/>
[M-C6H10O5-H] <sup>-</sup>	1	-163.06064845057	<input type="checkbox"/>
[M-C6H8O6-H] <sup>-</sup>	1	-177.03991300599	<input type="checkbox"/>
[M+CH3COONa-H] <sup>-</sup>	1	80.99524996793	<input type="checkbox"/>
[2M-H] <sup>-</sup>	1	-1.00782503207	<input checked="" type="checkbox"/>
[2M+FA-H] <sup>-</sup>	1	44.99765396793	<input checked="" type="checkbox"/>
[2M+Hac-H] <sup>-</sup>	1	59.01330396793	<input type="checkbox"/>
[3M-H] <sup>-</sup>	1	-1.00782503207	<input type="checkbox"/>
[M-2H]2 <sup>-</sup>	2	-2.01565006414	<input type="checkbox"/>
[M-3H]3 <sup>-</sup>	3	-3.02347509621	<input type="checkbox"/>

Load ☒ Together with Alignment Finish Cancel

User-defined adduct setting

Adduct: [M-H2O-H]<sup>-</sup> Mass: -19.0184 Xmer: 1 Charge: 1 Ion mode: Negative Add

You can select the adduct ions suitable. If necessary, you can add your own adduct ion by clicking “**User-defined adduct**” or load a whole adduct list by clicking “**Load**”.

## f. Alignment parameters

**Analysis parameter setting**

Alignment parameters setting

Result name: alignmentResult\_2020\_4\_14\_11\_47\_44

Reference file: QC-T3-neg3

Retention time tolerance: 0.1 min

MS1 tolerance: 0.015 Da

Advanced

Retention time factor: 0.5 (0-1)

MS1 factor: 0.5 (0-1)

Peak count filter: 0 %

N% detected in at least one group: 0 %

Remove features based on blank information: ☐

Sample max / blank average: 5 fold change

Keep 'reference matched' metabolite features: ☒

Keep 'suggested (w/o MS2)' metabolite features: ☐

Keep removable features and assign the tag: ☒

Gap filling by compulsion: ☐

Load ☒ Together with Alignment Finish Cancel

Specify the QC file. All samples will be aligned on this QC file.

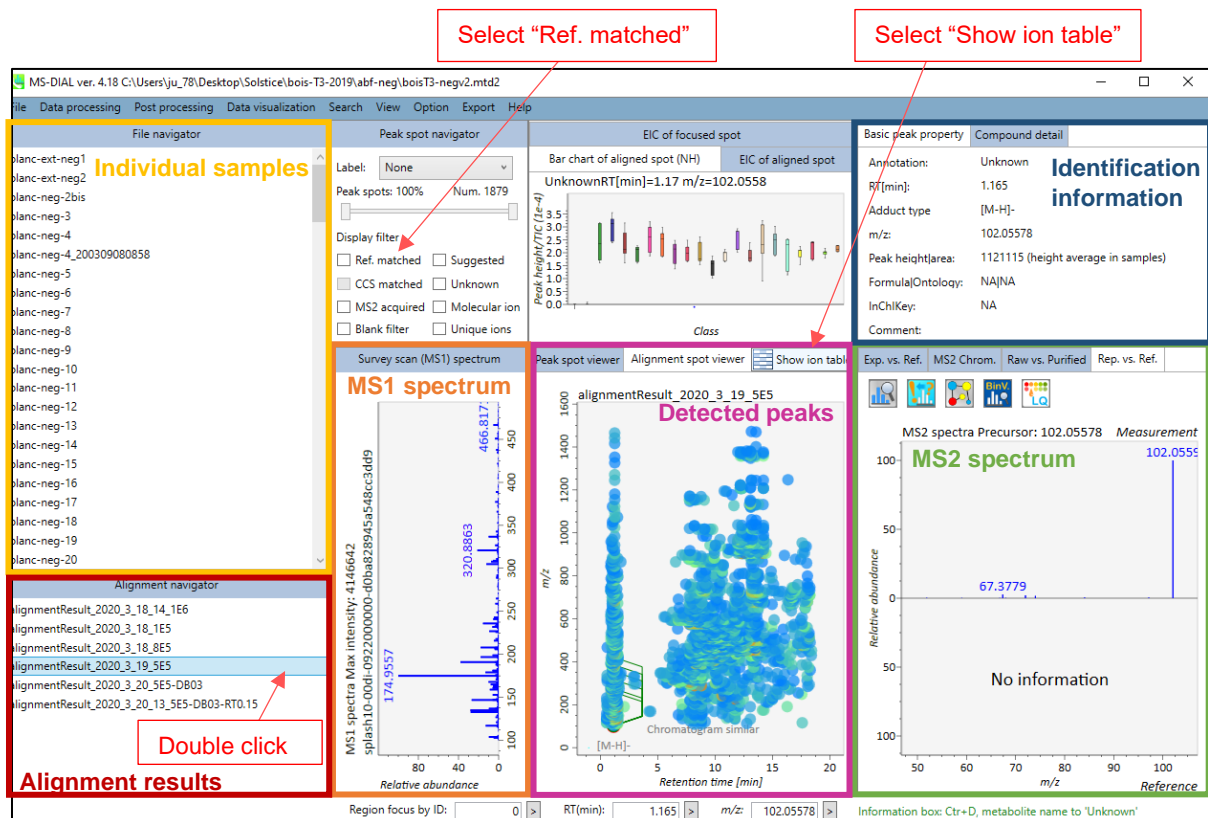
Set the retention time and mass tolerances

To remove peaks with more than N% missing values.

Peaks should be detected in N% of all samples of a class

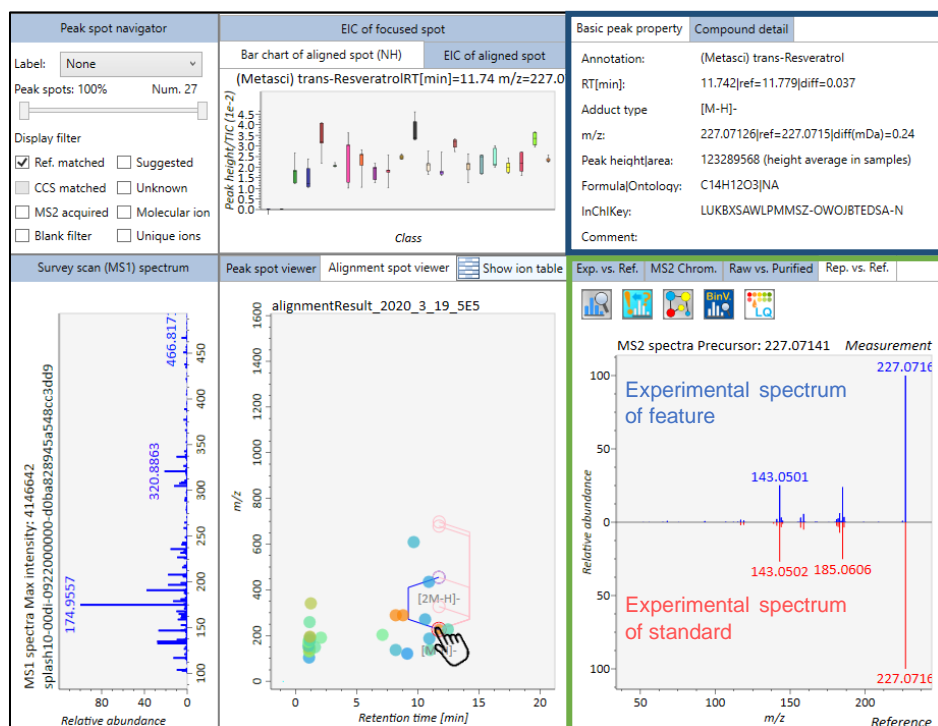
## 4. Analysis results

The following window appears after processing:





By clicking on “**Ref. matched**”, you can see the number of matched features based on your imported database. The following example is based on the interrogation of a msp format database based on retention time, accurate mass and MS/MS spectrum for identification of level 1:



By clicking on “**Show ion table**”, you will easily see all the identified features and their respective intensities in each defined class:

Alignment Table

Num of rows: 27

Metabolite Name Filter

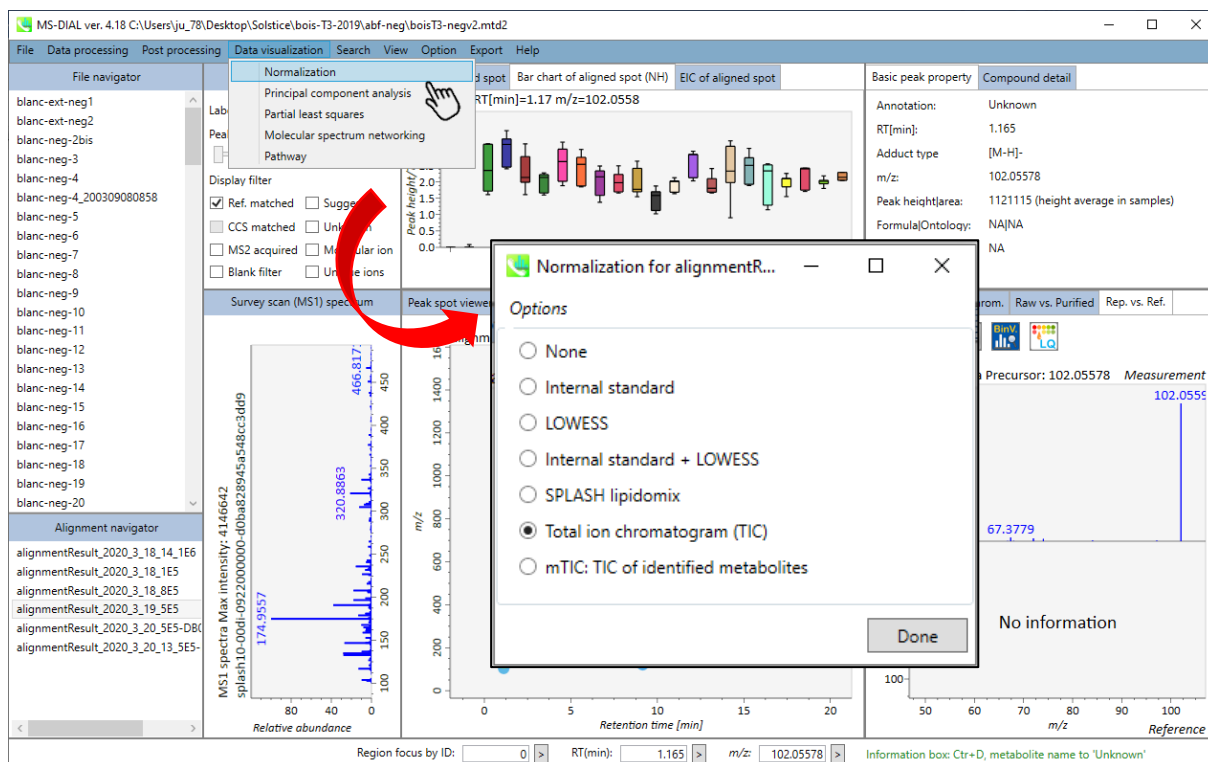
Comment Filter

102.06 Mz Range 1472.43 0.6 RT Range 18.4

ID	RT(min)	m/z	Type	Fill %	Metabolite name	Comment	Correlation	S/N	ANOVA P-value	Fold change (Max/Min)	BarChart
1	1.12	104.0351	[M-H] <sup>-</sup>	0.09	(IROA) SERINE		0.65	131.5	5.38E-05	118.12	
453	1.26	341.1085	[M-H] <sup>-</sup>	0.84	(IROA) MELIBIOSE		0.51	6289.5	7.17E-03	769.15	
310	8.21	289.0719	[M-H] <sup>-</sup>	0.84	(Metasci) (+)-Catechin (Hydrate)		0.67	29656.4	9.01E-09	20097.32	
309	8.81	289.0718	[M-H] <sup>-</sup>	0.84	(Metasci) (-)-Epicatechin		0.67	28564.6	3.13E-09	15026.86	
264	10.62	271.0614	[M-H] <sup>-</sup>	0.84	(Metasci) (+/-)-Naringenin		0.77	42.9	3.48E-13	1036319.00	
239	1.16	259.0224	[M-H] <sup>-</sup>	0.84	(IROA) MANNOSE 6-PHOSPHATE		0.80	1069.7	1.28E-15	475.27	
189	12.48	227.0714	[M-H] <sup>-</sup>	0.83	(Metasci) cis-Resveratrol		0.63	96.2	3.37E-02	5959.37	
185	11.74	227.0713	[M-H] <sup>-</sup>	0.84	(Metasci) trans-Resveratrol		1.00	9416.3	4.66E-15	20092.79	
146	7.13	203.0825	[M-H] <sup>-</sup>	0.83	(IROA) TRYPTOPHAN		0.39	1447.9	1.11E-03	3711.93	

After alignment process, on the “**data visualization**” menu select “**normalization**” and normalized data by Total ion chromatogram (TIC) or another method.



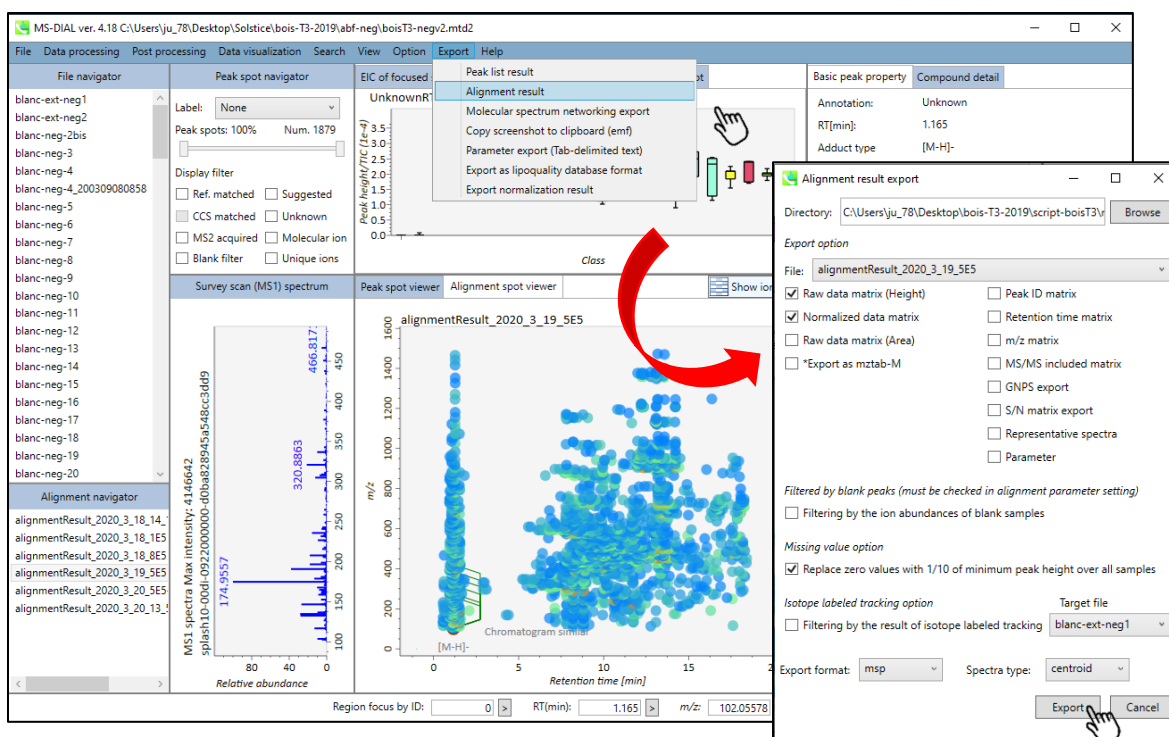


On the “**data visualization**”, after normalization, you will be able to do some statistical analysis such as Principal component analysis.

## 5. Export peak list

On the “**Export**” menu select “**alignment results**” and export both **Raw data matrix (Height)** and **Normalized data matrix**. You can also select the option “**Replace zero values with 1/10 of minimum peak height over all samples**” for further statistical purpose.





## 6. Export all peaks

By right-clicking on one point, on the **“search formula and structure”** menu select **“Add component to search list”** and export **« all peaks »** for further identification purpose with MS-FINDER (see MS-FINDER tutorial). On the chosen folder, .MAT files will be generated and directly importable into MS-FINDER. At this point, you should be ready to follow the MS-CleanR and MS-FINDER tutorials.

