Metabolomics Software Tools

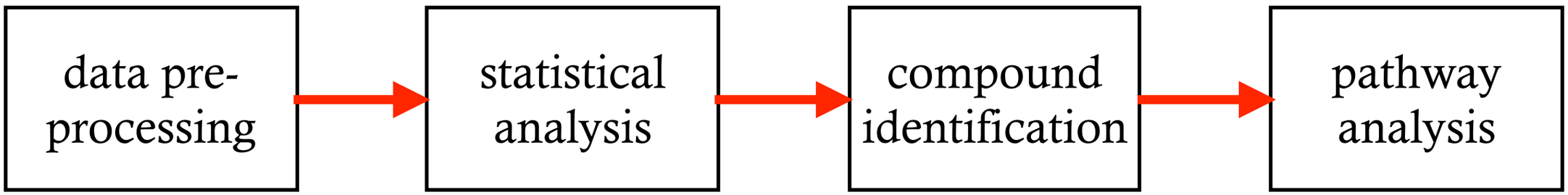
Xiuxia Du, Paul Benton, Stephen Barnes

Outline

- **Introduction**
- **Software Tools for LC-MS metabolomics**
- **Software Tools for GC-MS metabolomics**
- **Software Tools for Statistical Analysis**

Introduction

- **LC-MS data analysis workflow**



- **GC-MS data analysis workflow**

Software Tools for LC-MS

- **Commercial**
  - MassHunter — Agilent
  - MetQuest — Thermo Scientific
  - MetWorks — Thermo Scientific
  - Multiple Mass Defect Filter — Thermo Scientific
  - Progenesis QI — Waters
  - XCMS^{plus} — Sciex

- and more ……

Software Tools for LC-MS

- **Free**
  - **Insilicos Viewer**

  - **ProteoWizard**
    - **SeeMS:** interactive viewer for mass spec data files (Windows only)
    - **MSConvert:** convert between various file formats

  - **MZmine 2:** LC-MS data processing

  - **MetaboSearch:** perform mass-based metabolite search simultaneously against four major metabolite databases

  - **MetaboAnalyst:** a web server for metabolomics data analysis

  - and more ……

Data format

- **Proprietary data formats:** need vendor library functions to access

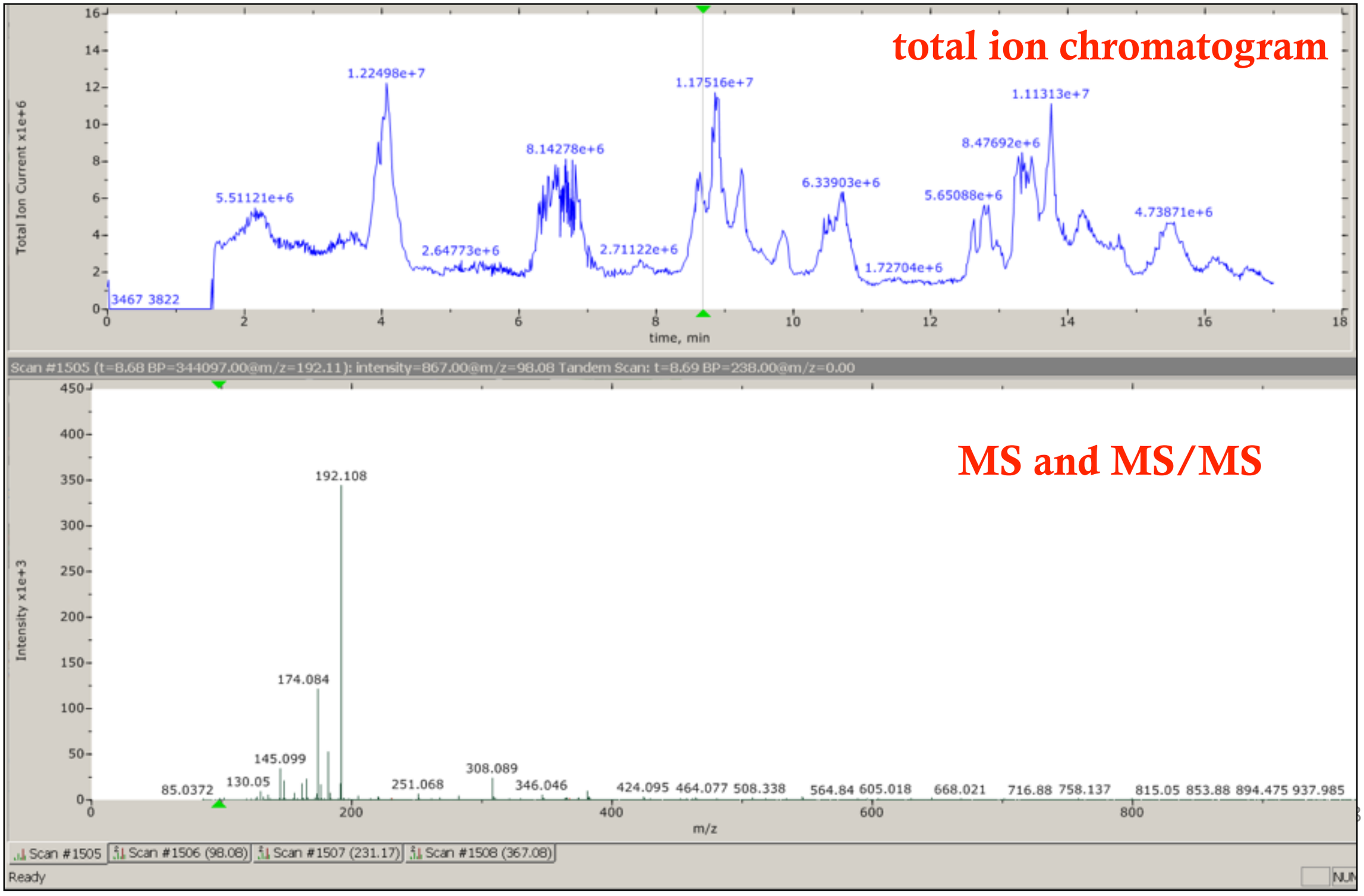
Company	File extension
Agilent	.D
Sciex	.WIFF
Theomo	.RAW
Waters	.RAW
……	

- **Open data formats:** free library functions to access
  - netCDF
  - mzData
  - mzXML
  - mzML

InsilicosViewer

Insilicos Viewer

Raw data viewer



ProteoWizard

ProteoWizard: SeeMS

- **A data viewer**

- **Can read these open data formats**
  - mzML
  - mzXML

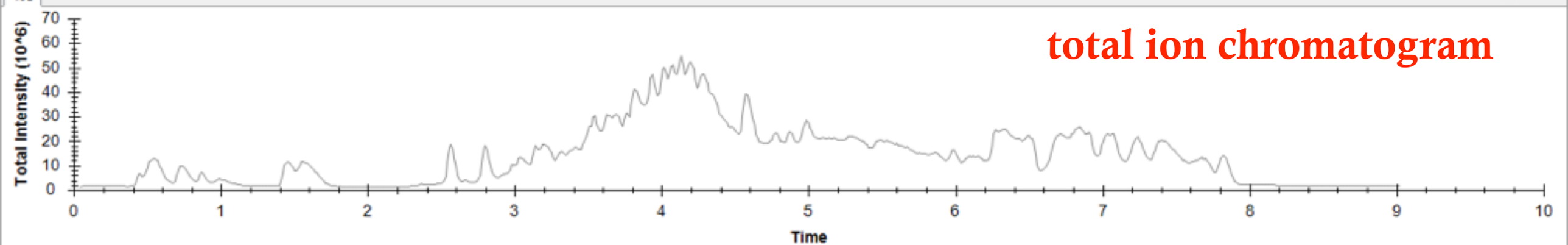
ProteoWizard: SeeMS

Raw data viewer

SeeMS

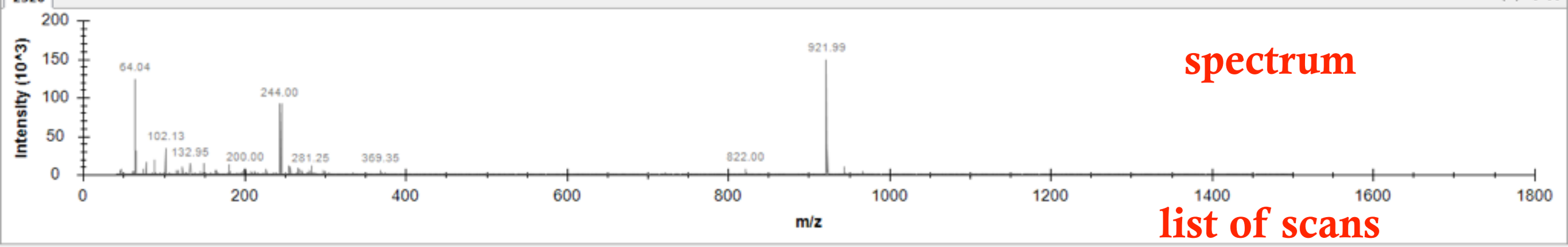
File View Window Help

TIC



total ion chromatogram

2320



spectrum

list of scans

D20100526-LC2-PP0000441-B4-I1-P spectra

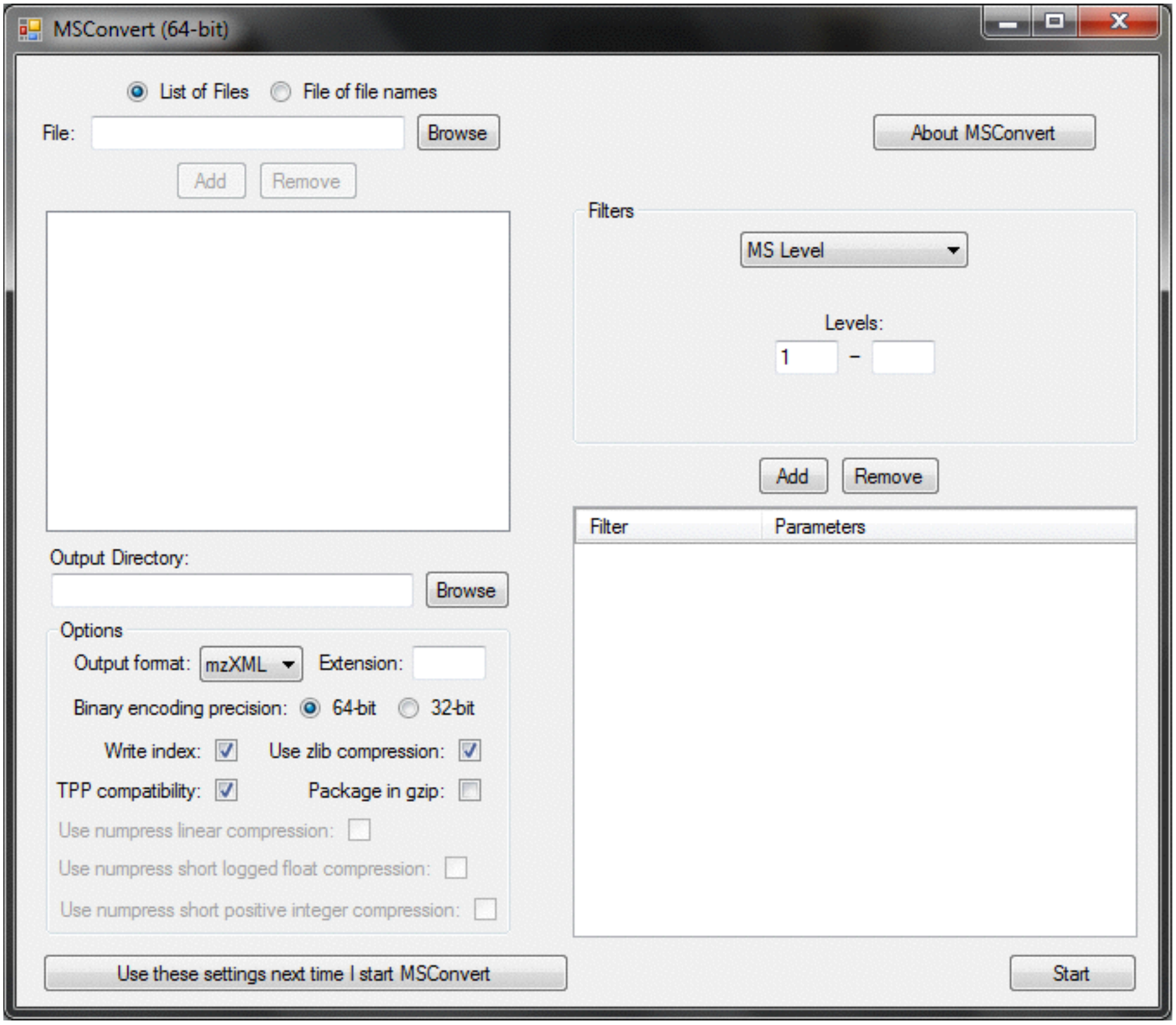
Scanid	Spectrum Type	MS Level	Data Points	Scan Time	Base Peak m/z	Base Peak Intensity	Total Ion Current	IC Id	DP Id	Precursor Info	Scan Info
2320	MS1 spectrum	1	113059	0.03855	922.014188022825	150386.094	1377161.5	IC1	unknown	n/a	[39.997786023848-...
3028	MS1 spectrum	1	113055	0.05035	922.013290199498	156199.609	1361376.75	IC1	unknown	n/a	[39.997786023848-...
3738	MS1 spectrum	1	113045	0.062166666667	922.014193573584	158492.219	1431714.63	IC1	unknown	n/a	[39.997786023848-...
4446	MS1 spectrum	1	113061	0.072966666667	922.013997182642	159991.484	1524202.75	IC1	unknown	n/a	[39.997786023848-...

D20100526-LC2-PP0000441-B4-I1-P chromatograms | D20100526-LC2-PP0000441-B4-I1-P spectra

Finished loading source metadata.

ProteoWizard: MSConvert

Data format converter



ProteoWizard: MSConvert

- **Supported data formats**
  - Read: open formats, vendor formats
  - Write: open formats

- **Filters and transformation**
  - msLevel
  - Peak picking
  - Zero samples
  - ETD filter
  - Threshold peak filter
  - Charge state predictor
  - Activation
  - Subset

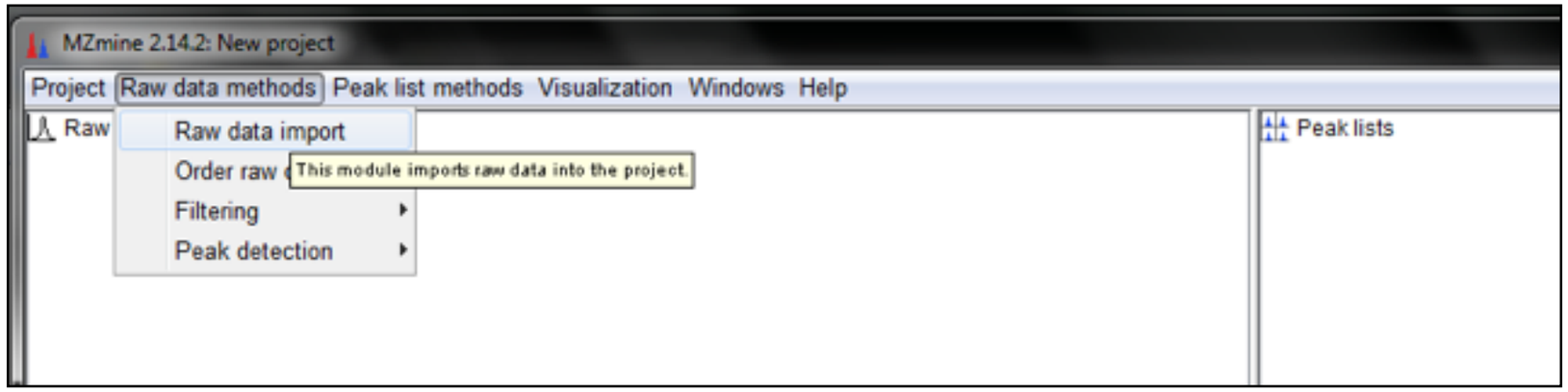
MZmine 2

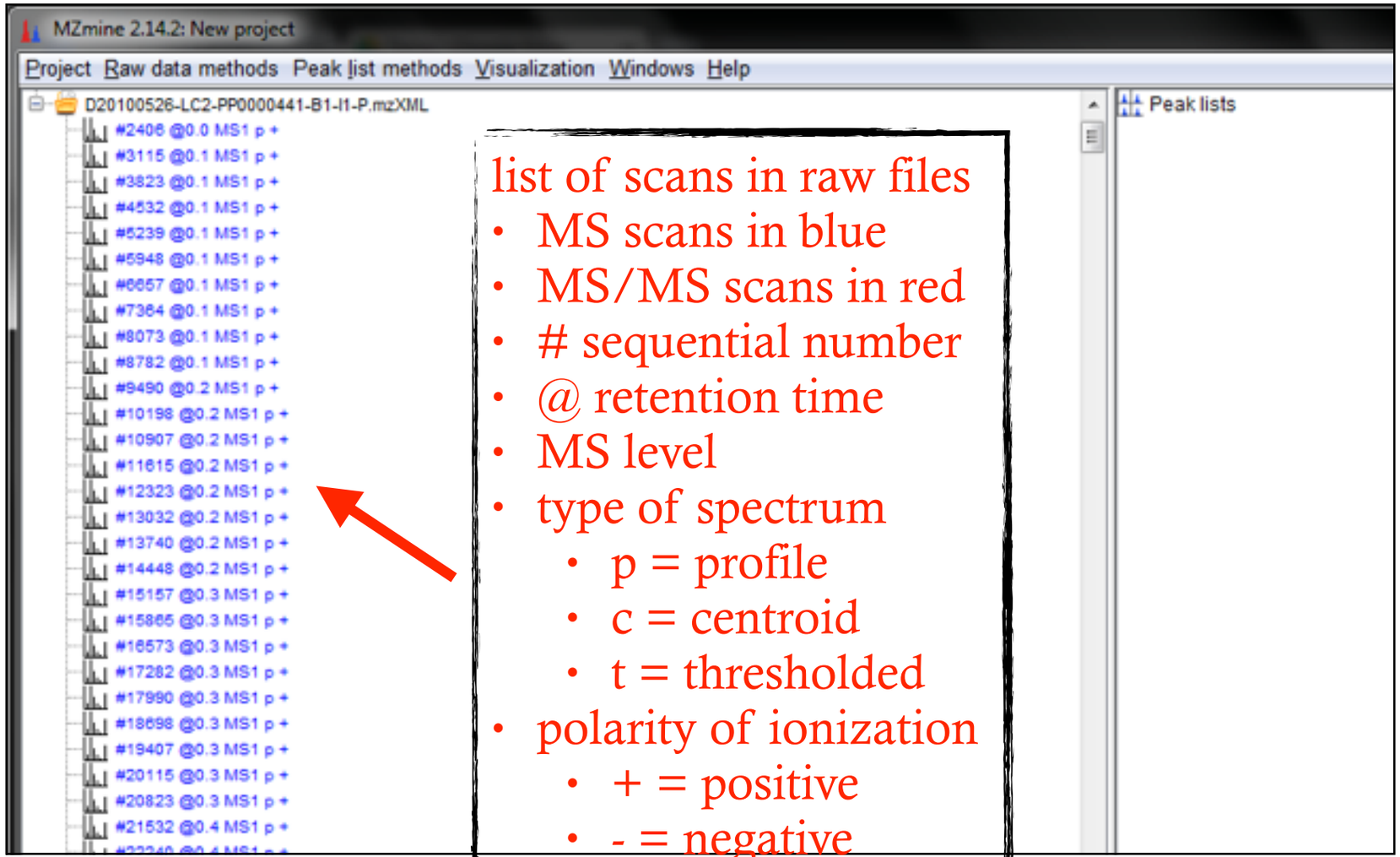
MZmine 2

- **LC-MS metabolomics data processing, analysis, and visualization**

- **Supported open data formats**
  - NetCDF
  - mzData
  - mzML
  - mzXML

Raw data import

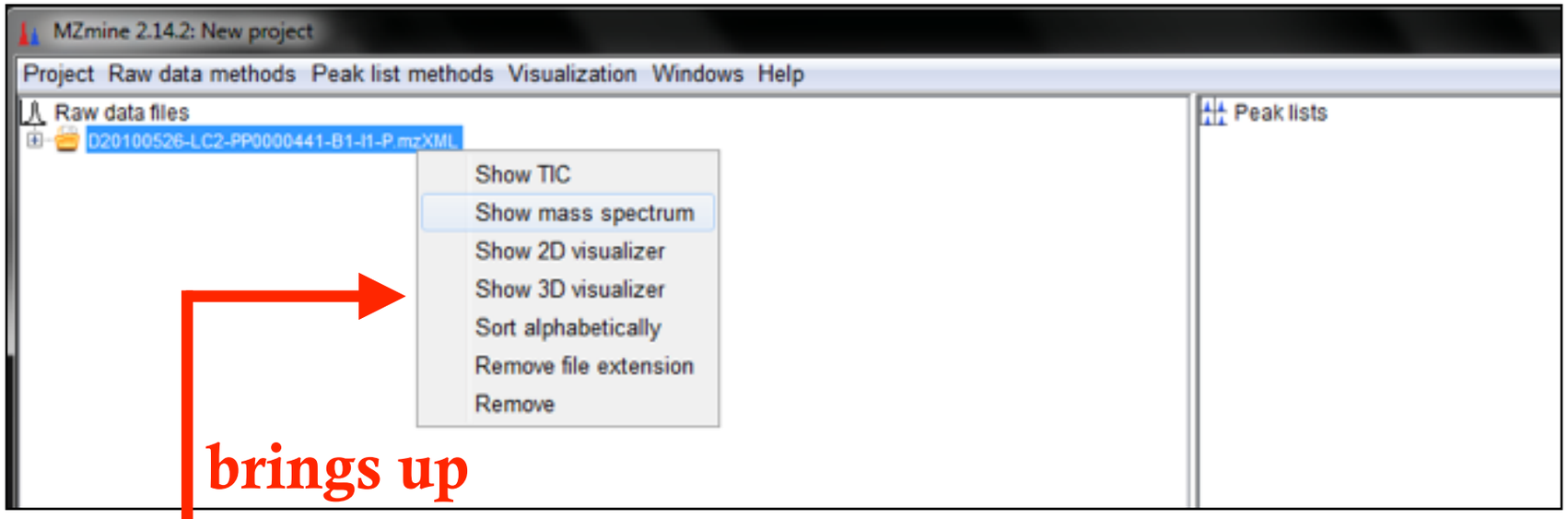




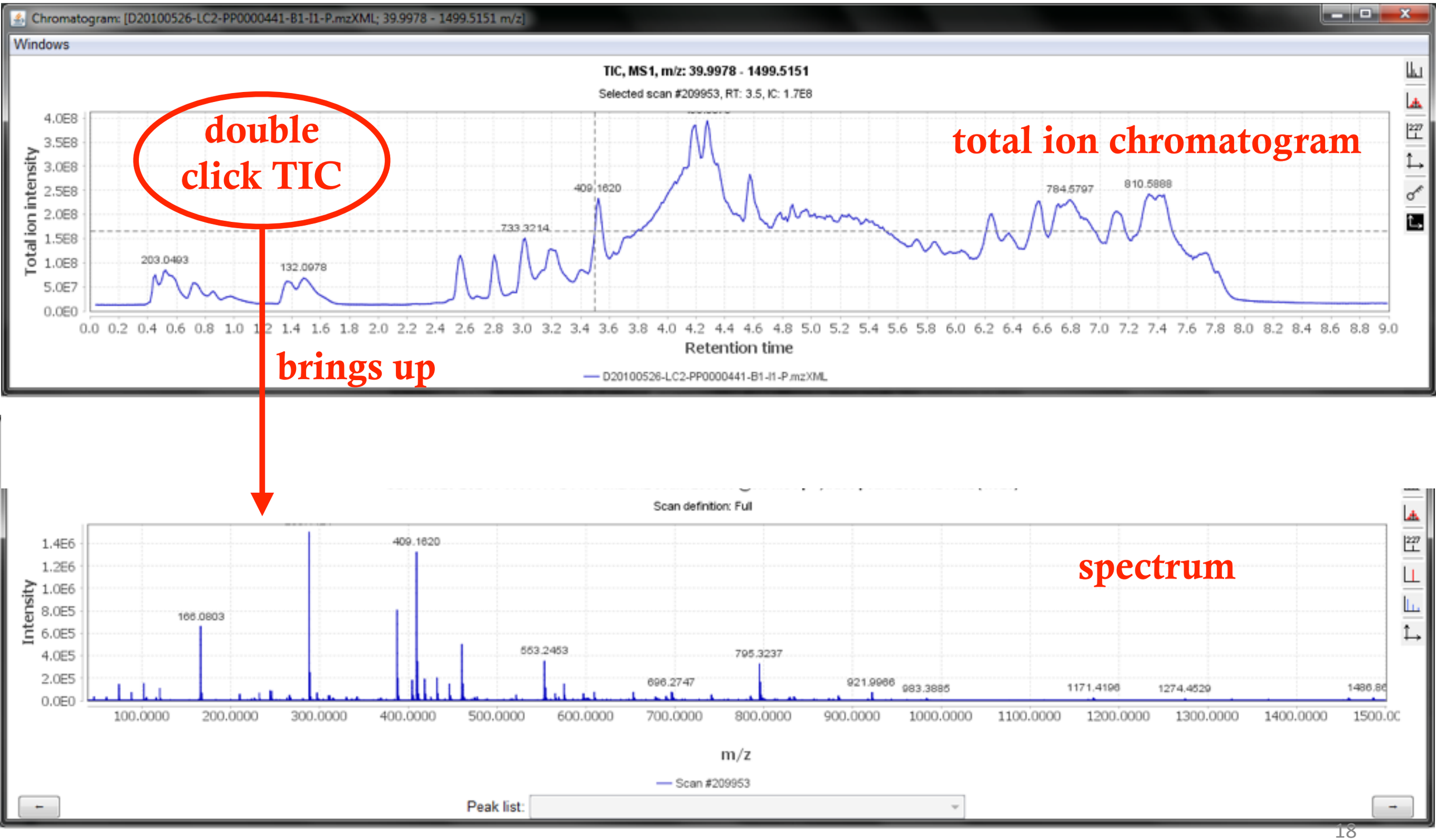
list of scans in raw files

- MS scans in blue
- MS/MS scans in red
- # sequential number
- @ retention time
- MS level
- type of spectrum
  - p = profile
  - c = centroid
  - t = thresholded
- polarity of ionization
  - + = positive
  - - = negative
  - ? = unknown

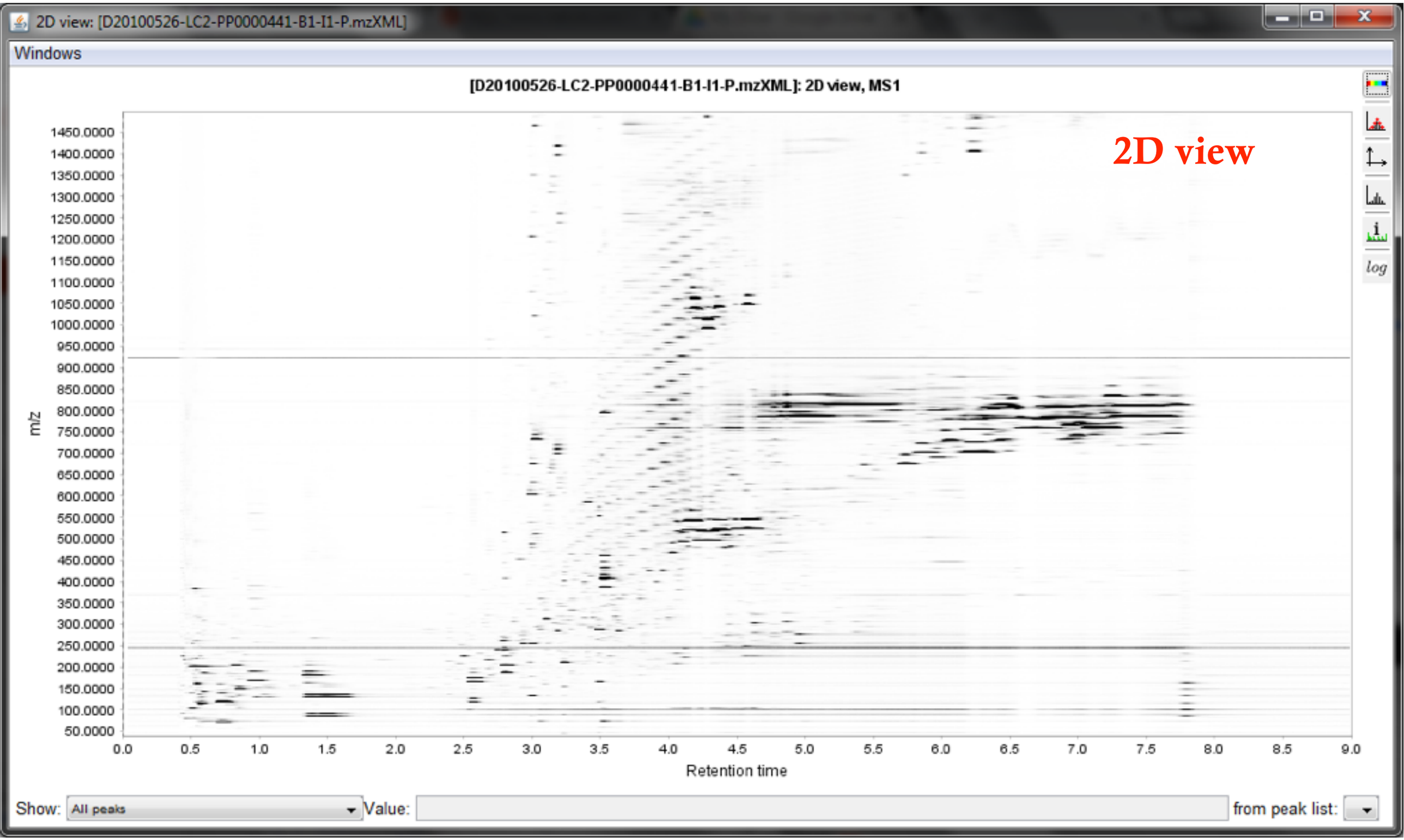
Raw data visualization



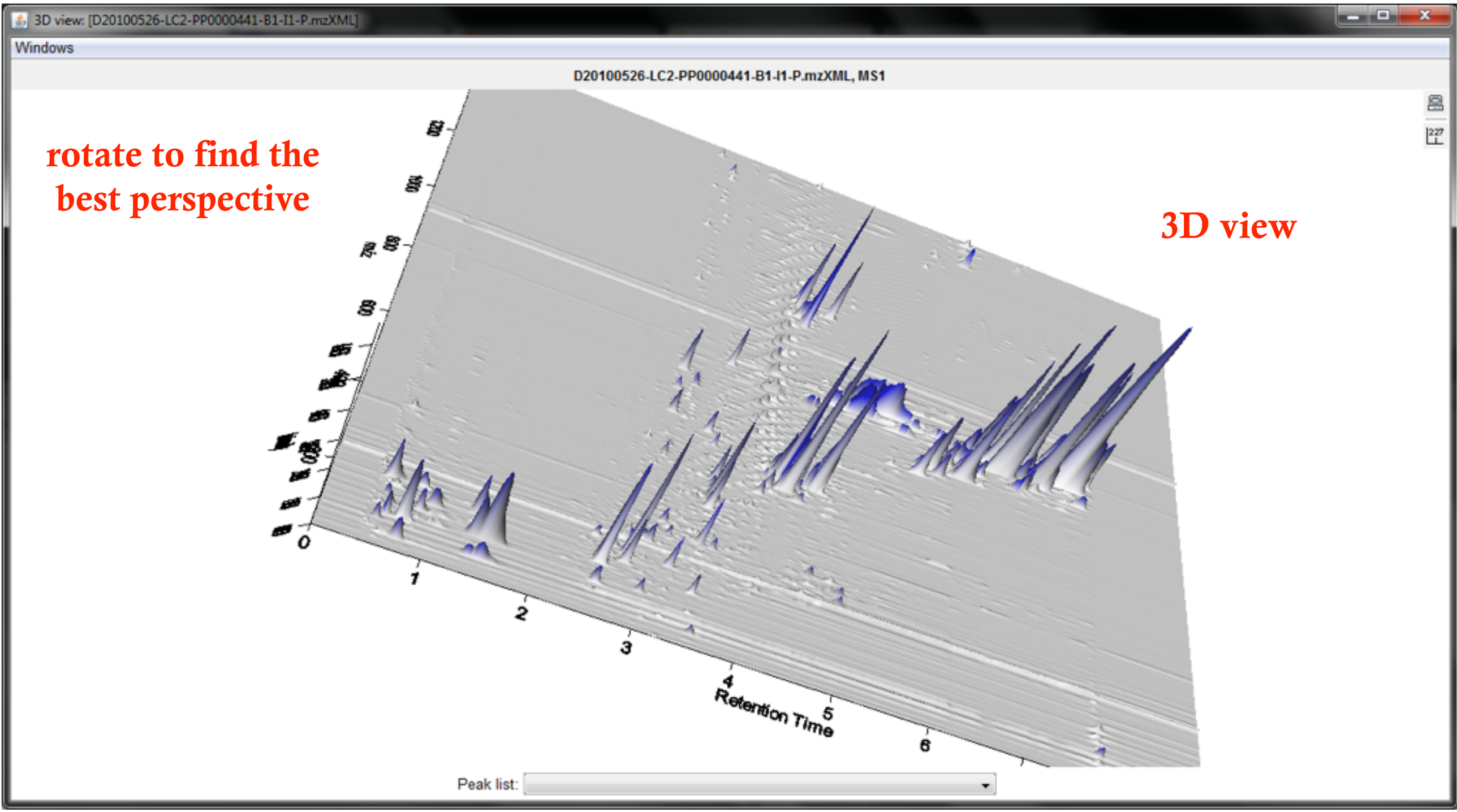
Raw data visualization



Raw data visualization



Raw data visualization



Data pocessing

- **Workflow**
  - Raw data import
  - (optional) Raw data methods / Filtering
  - Peak detection
  - Isotopic peak grouping
  - (optional) Identification of fragments, adducts, and peak complexes
  - (optional) Normalization of retention time
  - Alignment
  - (optional) Gap filling
  - (optional) Normalization of peak heights / areas
  - (optional) Identification using database search, formula prediction, etc.
  - Data analysis and export

Peak detection

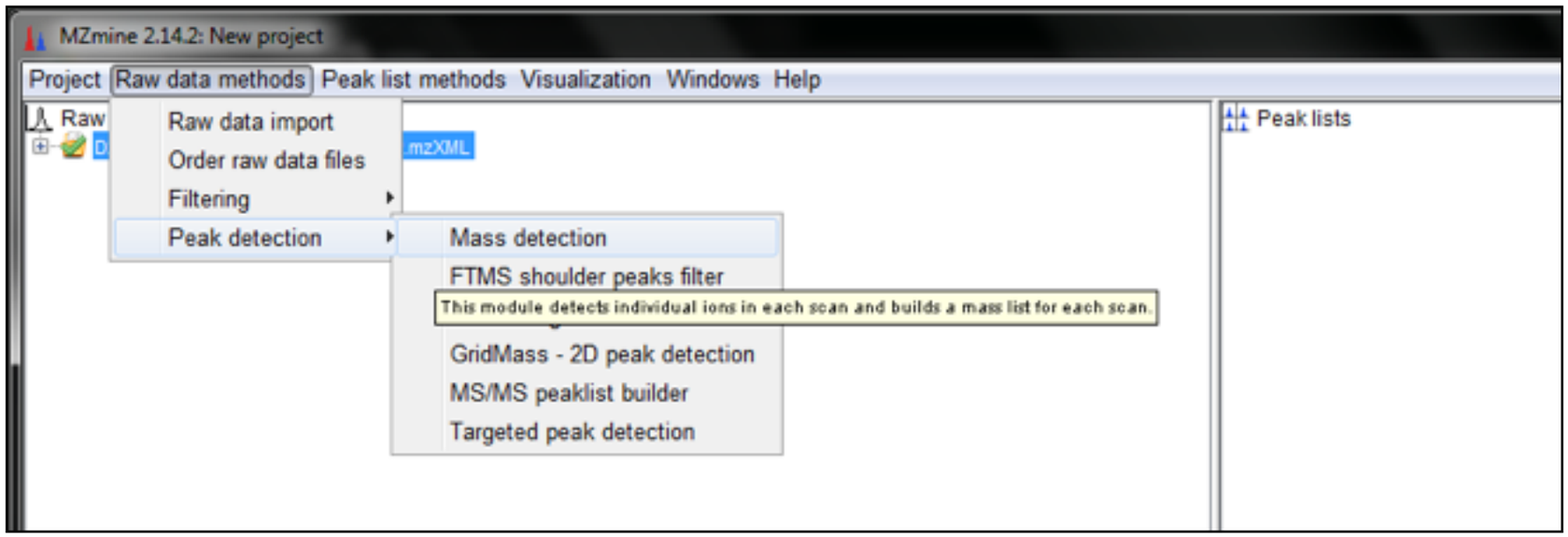
- **Mass detection**
  - Centroid
  - Exact mass
  - Local maxima
  - Recursive threshold
  - Wavelet transform

- **Chromatogram building**

- **Peak deconvolution**

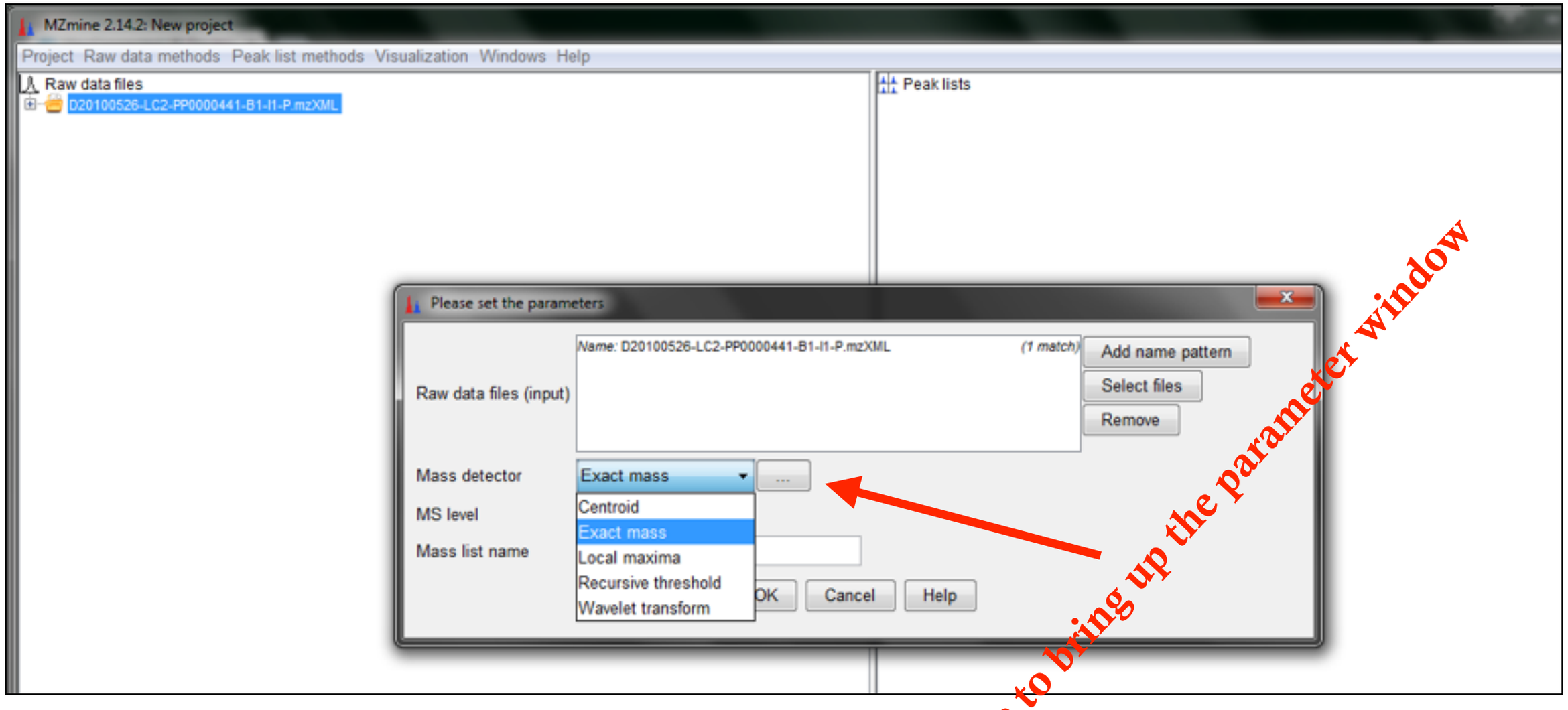
Peak detection

Peak detection procedure



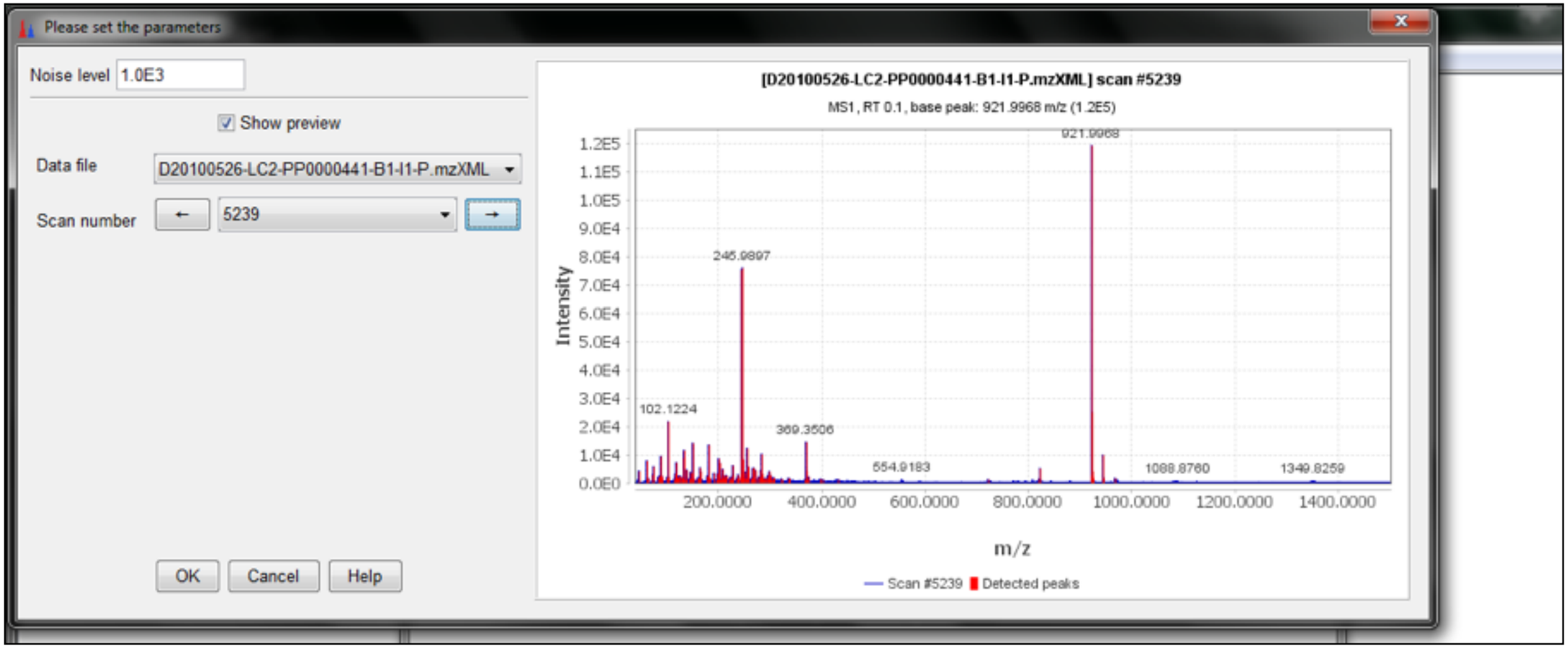
Mass detection

Mass detection options



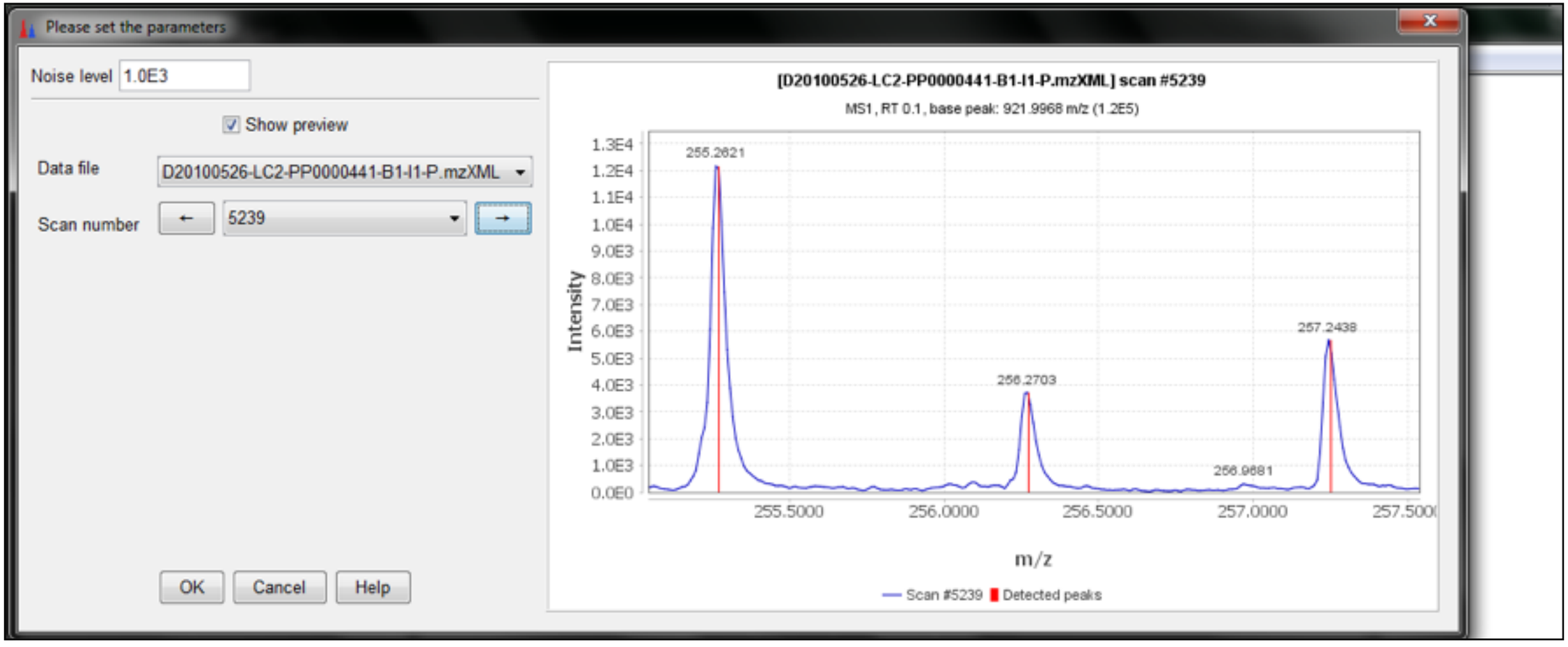
Mass detection

Set mass detection parameter



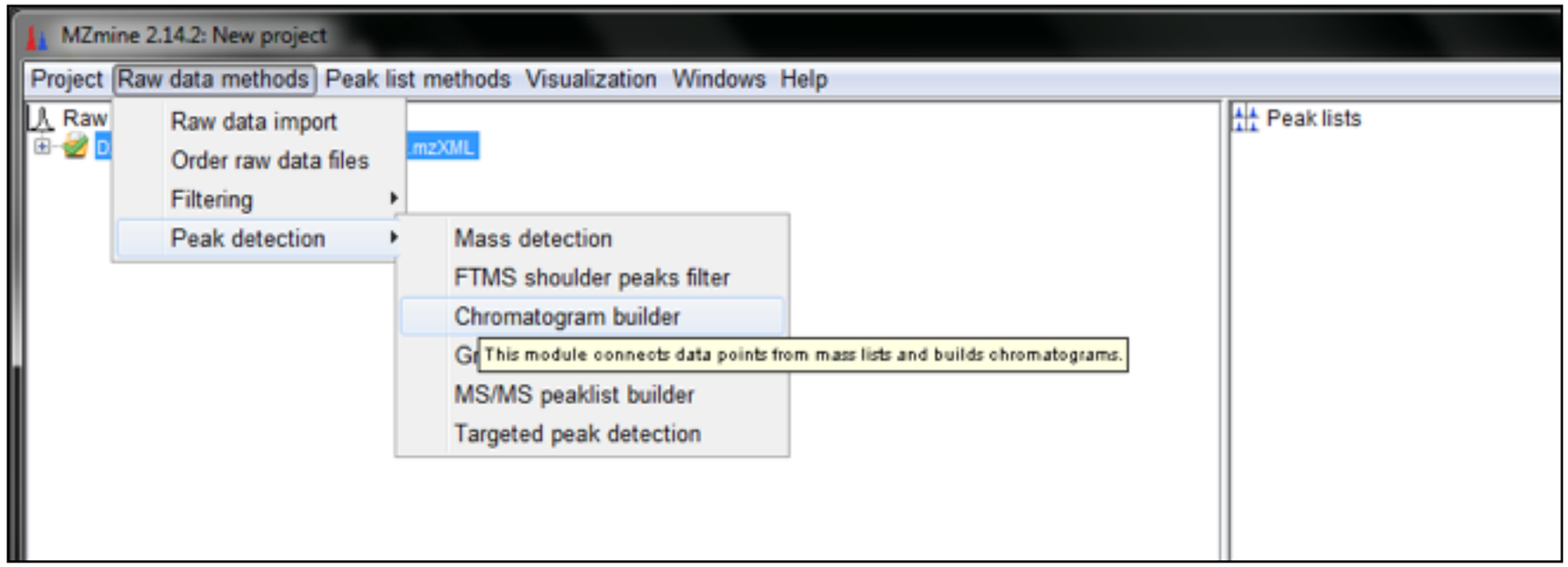
Mass detection

Set mass detection parameter



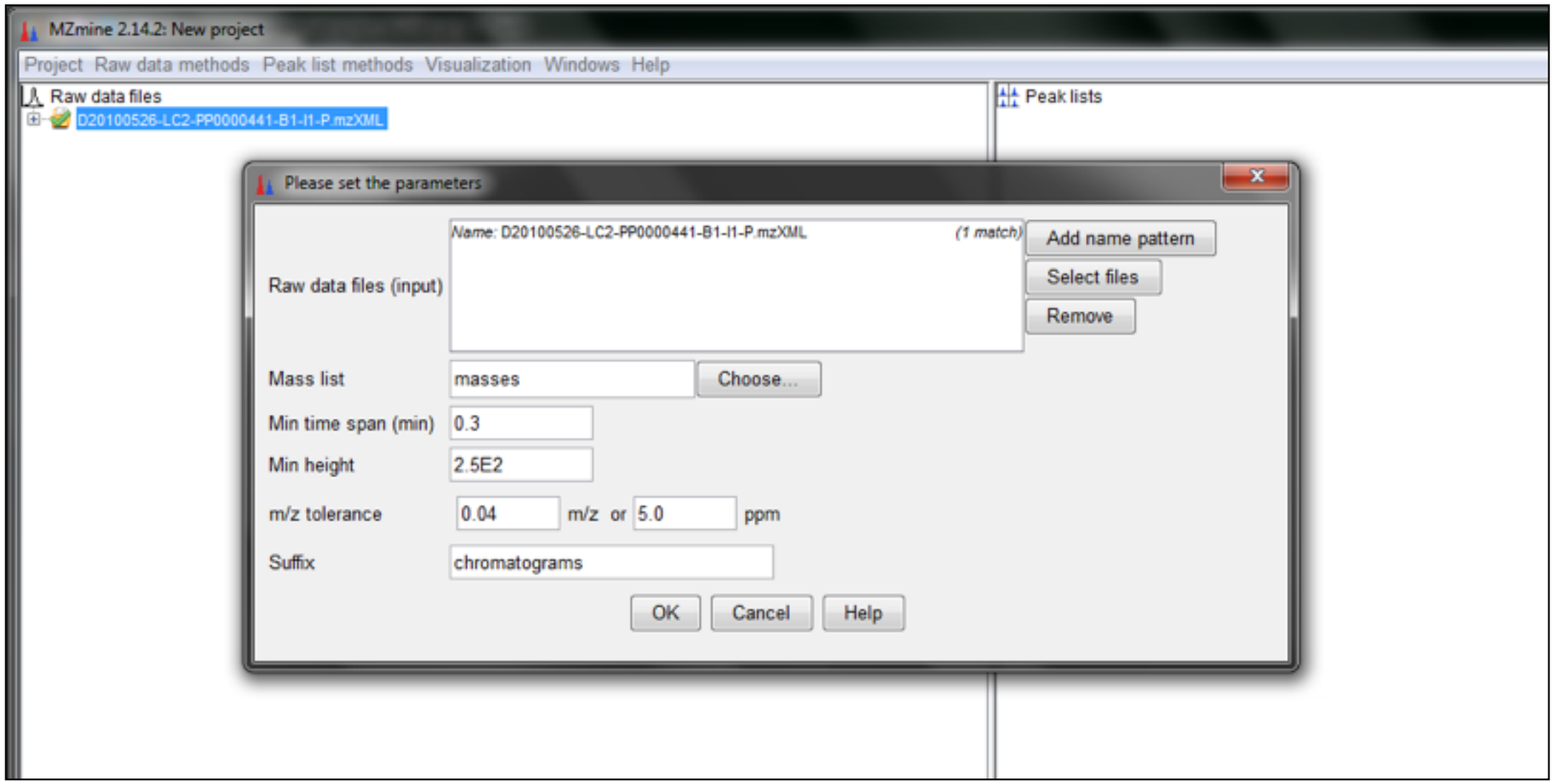
Chromatogram building

Chromatogram builder



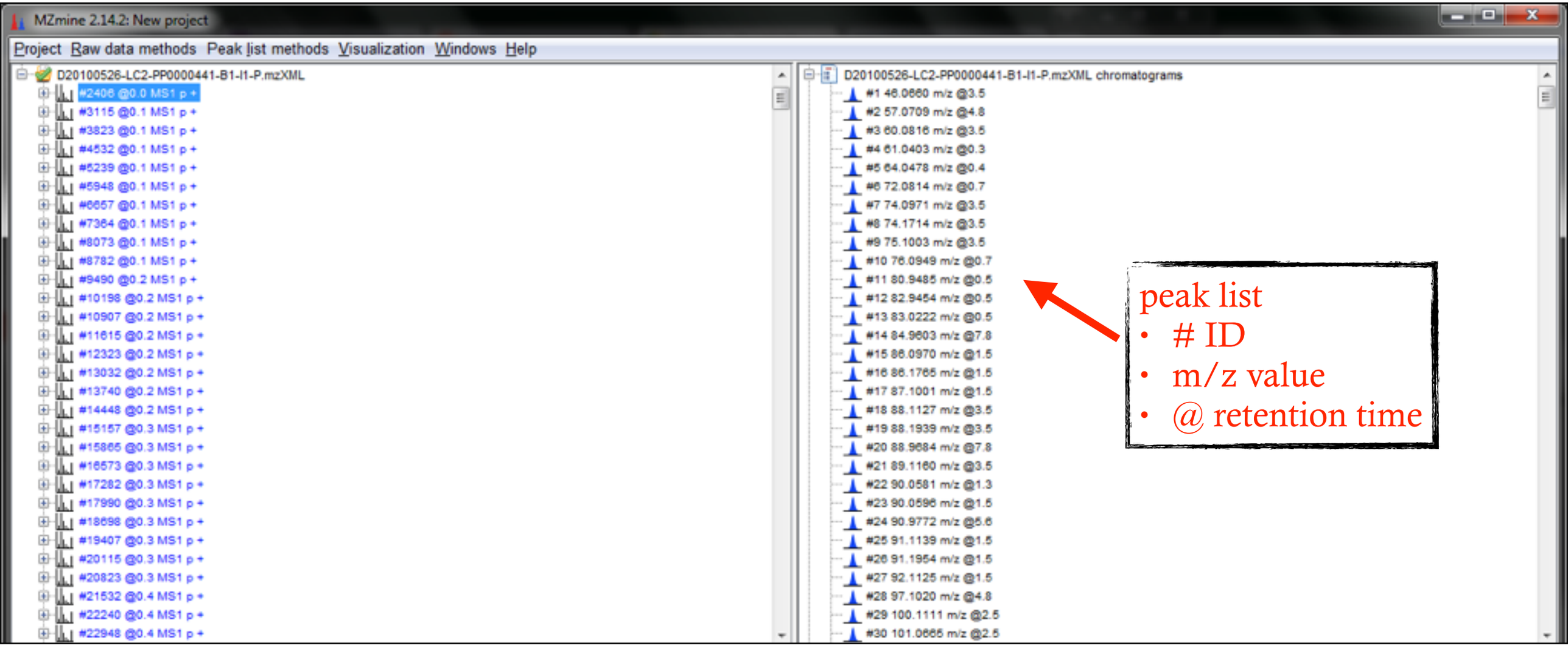
Chromatogram building

Chromatogram builder — set parameters



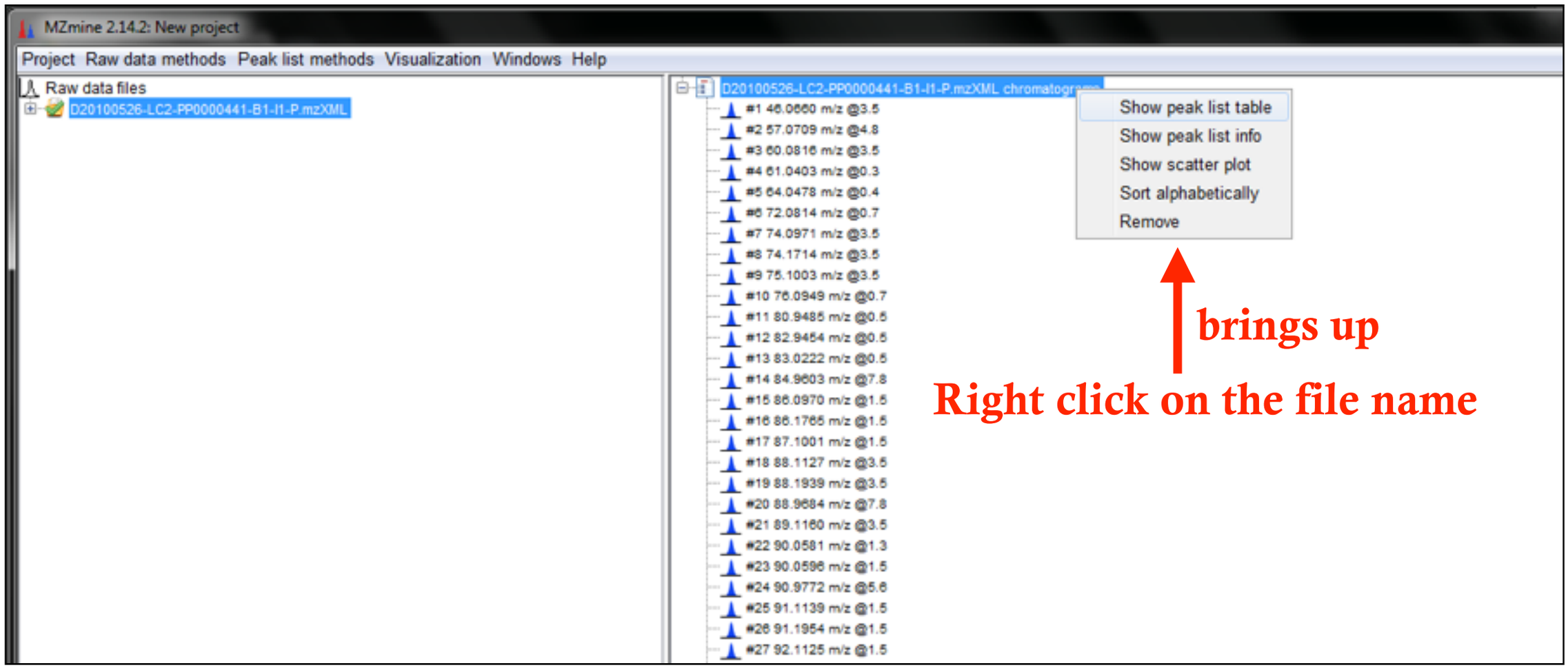
Chromatogram building

After chromatograms are built



Chromatogram building

To display extracted ion chromatograms



Chromatogram building

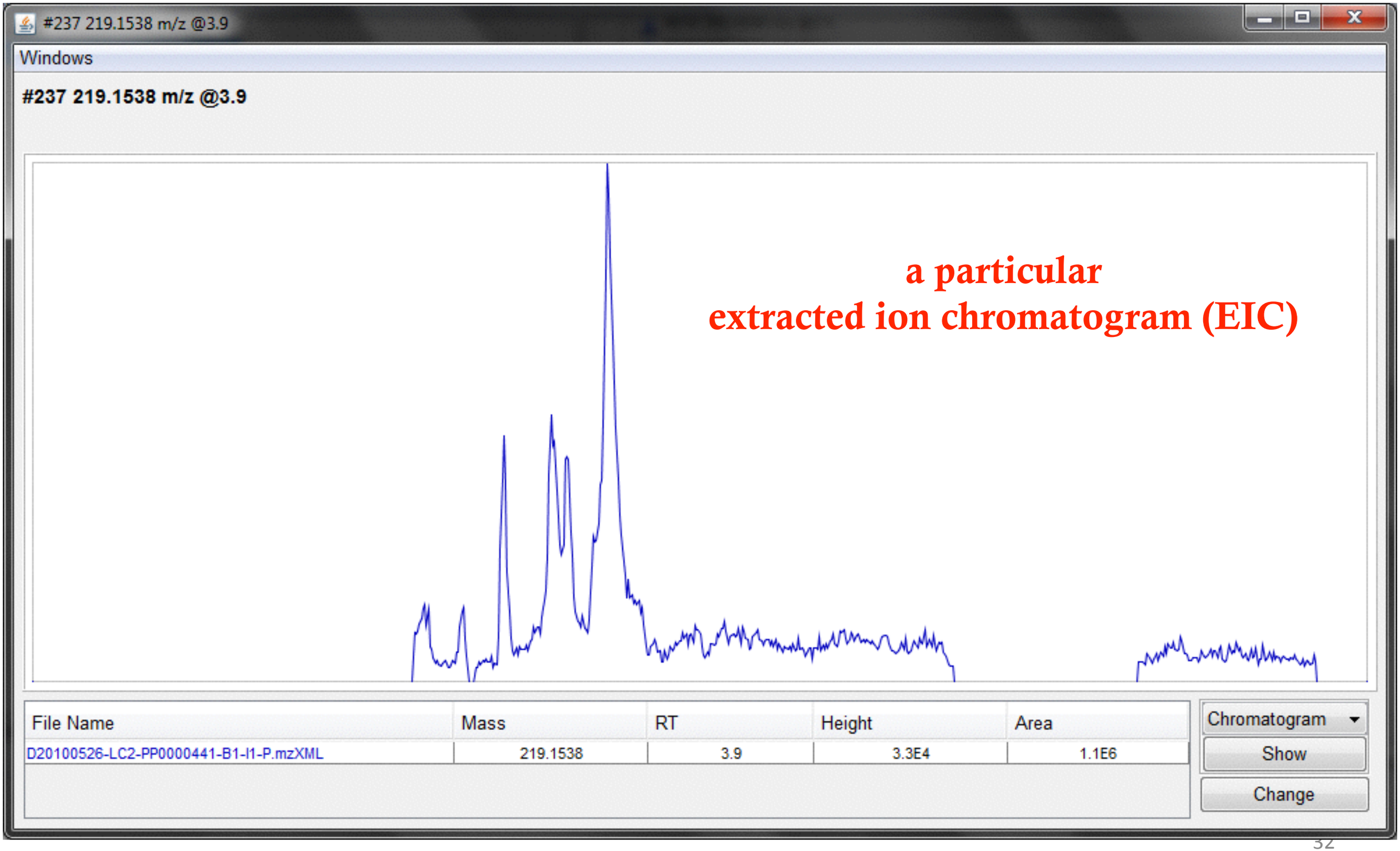
Show chromatogram information

Peak list: D20100526-LC2-PP0000441-B1-I1-P.mzXML chromatograms

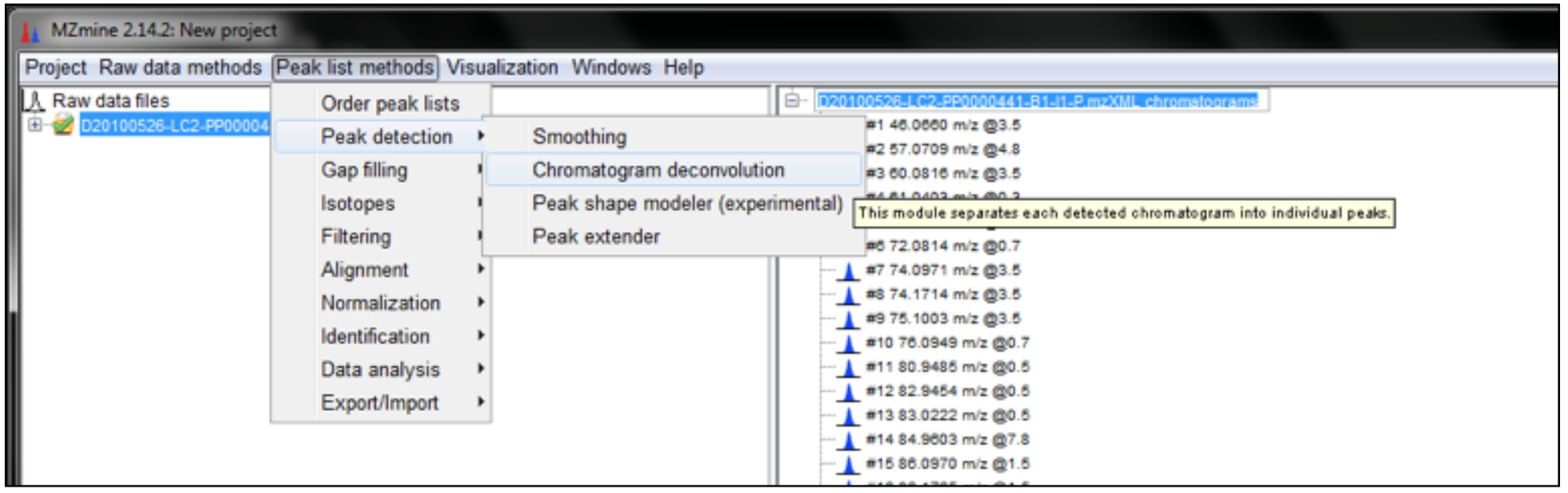
Windows

ID	Average		Identity	Comment	Peak shape	D20100526-LC2-PP0000441-B1-I1-P.mz...		
	m/z	RT				Status	Height	Area
227	213.1133	3.2					1.2E5	2.1E6
228	213.9883	0.3					3.7E3	6.0E4
229	213.9911	0.6					7.8E3	9.1E5
230	215.0170	0.9					5.2E4	5.0E5
231	215.1392	1.1					2.4E3	5.1E4
232	215.1558	3.8					8.9E3	6.4E4
233	215.9872	0.9					5.0E3	9.7E5
234	217.1555	2.7					2.4E5	2.2E6
235	218.1690	1.9					3.7E4	7.1E5
236	219.0270	0.5					6.9E4	4.2E5
237	219.1538	3.9					3.3E4	1.1E6
238	220.9361	7.7					6.5E3	4.2E5
239	221.0030	8.2					2.3E3	1.1E5
240	221.1832	3.2					4.2E4	1.0E6
241	222.0150	1.6					7.4E3	1.2E5
242	223.1308	3.4					3.0E4	1.9E5
243	223.2062	4.1					1.9E5	4.9E6
244	224.2097	4.1					3.3E4	7.3E5
245	225.1299	3.6					1.1E5	6.9E5
246	225.2126	4.5					6.2E3	6.1E5
247	226.1574	3.5					3.9E4	1.7E5

Chromatogram building

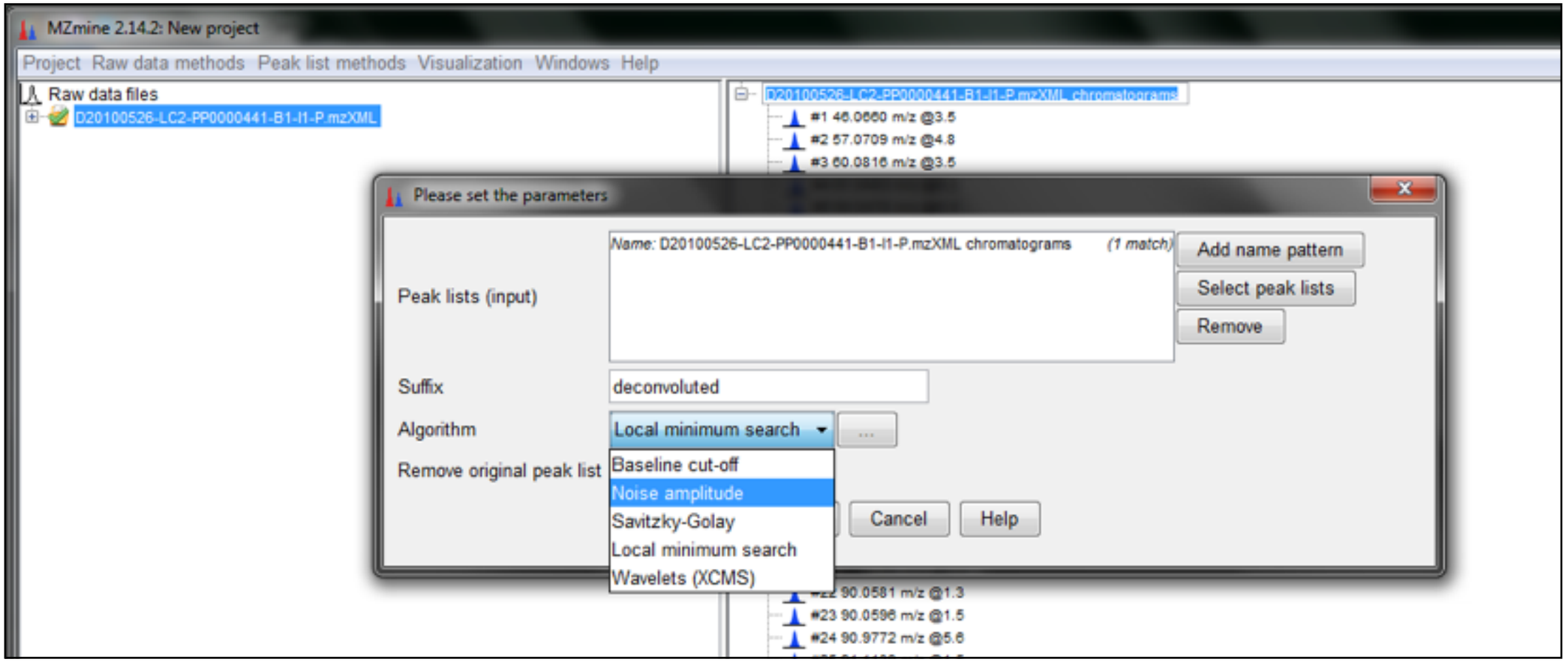


Peak deconvolution



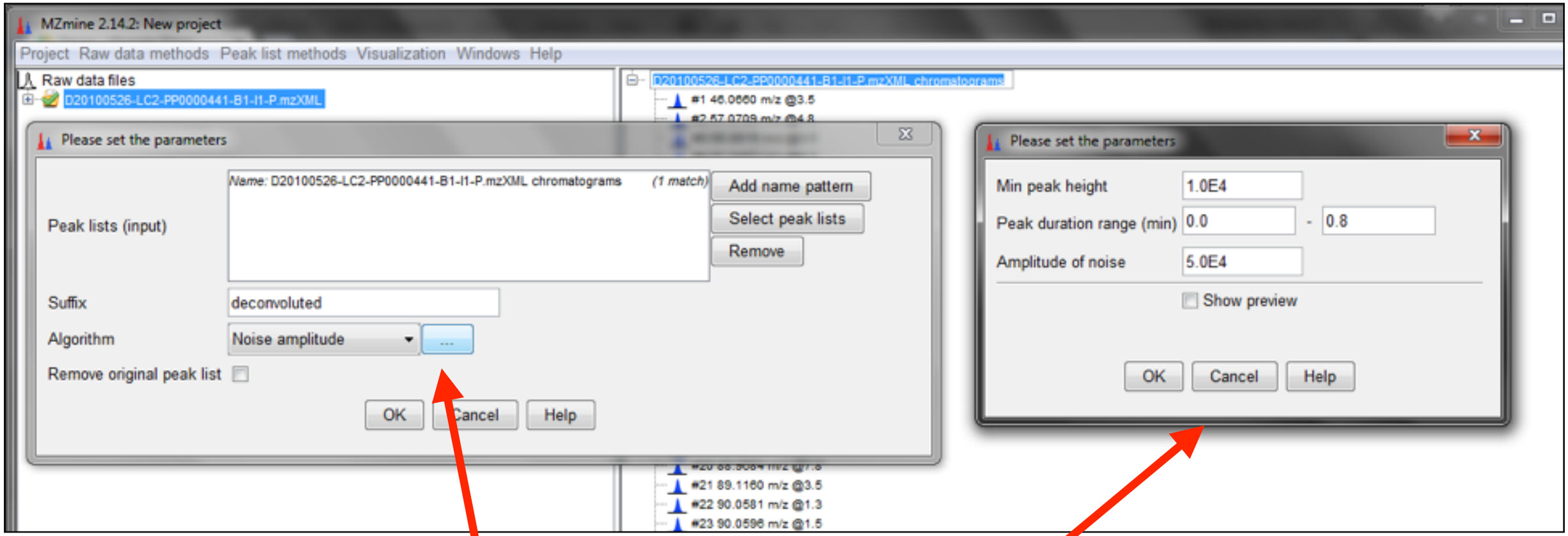
Peak deconvolution

Algorithms



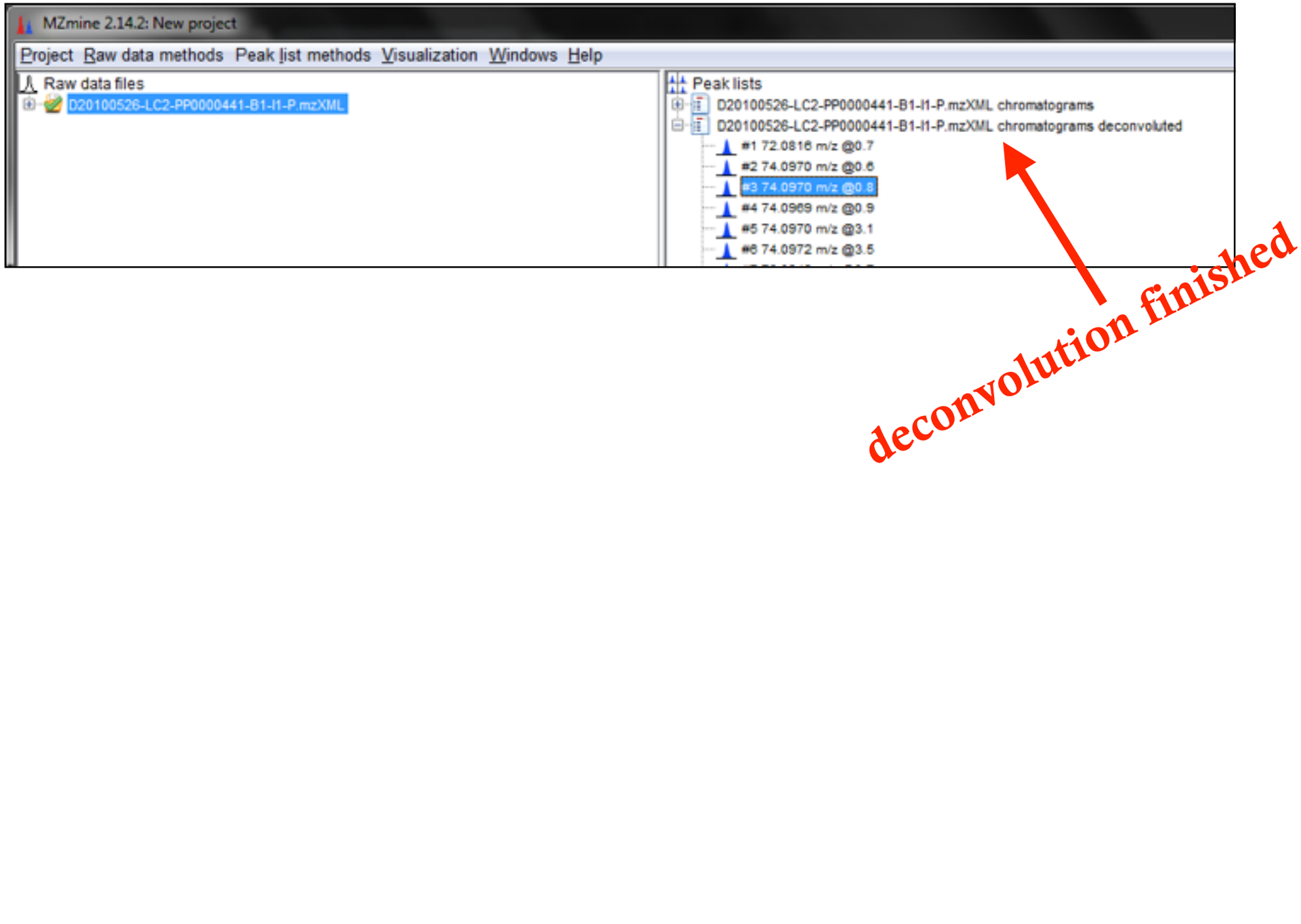
Peak deconvolution

Set parameters



click here to bring up the parameter window

Peak deconvolution



Peak deconvolution

Results

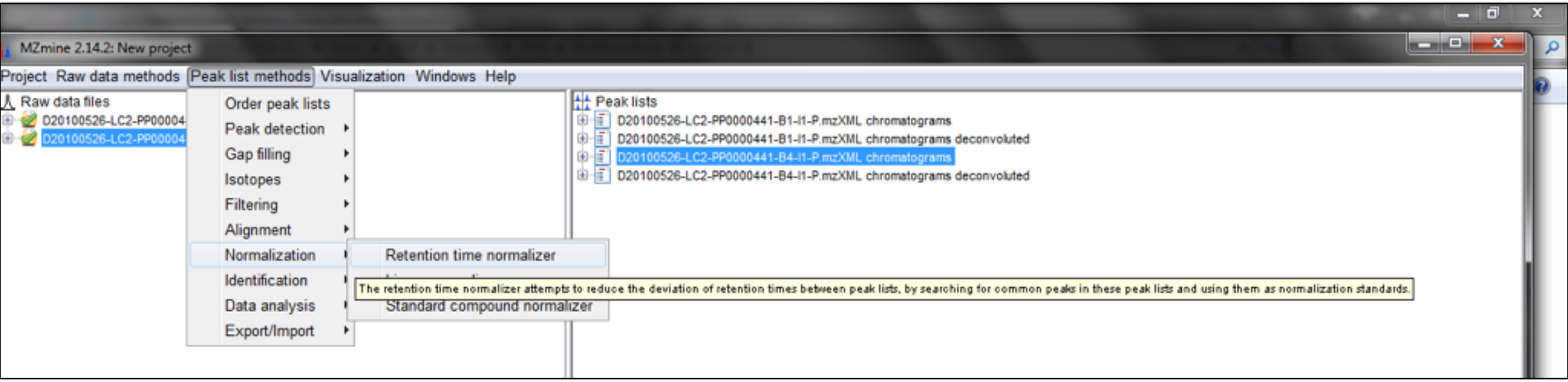
Peak list: D20100526-LC2-PP0000441-B1-I1-P.mzXML chromatograms deconvoluted

Windows

ID	Average		Identity	Comment	Peak shape	D20100526-LC2-PP0000441-B1-I1-P.mz...		
	m/z	RT				Status	Height	Area
85	187.0587	0.6					2.9E5	8.5E5
86	191.0183	0.9					3.1E5	1.9E6
87	191.1122	1.4					4.8E5	2.5E6
88	195.1237	2.7					1.2E5	3.1E5
89	203.0536	0.5					1.5E6	7.0E6
90	204.1238	0.8					5.2E5	2.0E6
91	204.0573	0.5					1.1E5	2.9E5
92	204.0638	1.4					1.2E5	5.7E5
93	205.0983	2.8					4.7E6	1.5E7
94	209.1909	4.1					1.6E5	6.5E5
95	210.0769	3.5					6.7E4	7.3E4
96	212.0909	3.2					1.3E6	3.5E6
97	213.0943	3.2					1.2E5	2.6E5
98	215.0168	0.9					5.2E4	7.1E4
99	217.1055	2.7					2.4E5	9.4E5
100	219.0275	0.5					6.9E4	2.0E5
101	223.2070	4.1					1.9E5	6.6E5
102	225.1497	3.6					1.1E5	2.4E5
103	227.1897	2.5					3.6E5	1.1E6
104	227.2382	4.6					1.7E5	1.9E6
105	229.1552	0.9					5.6E4	3.7E4

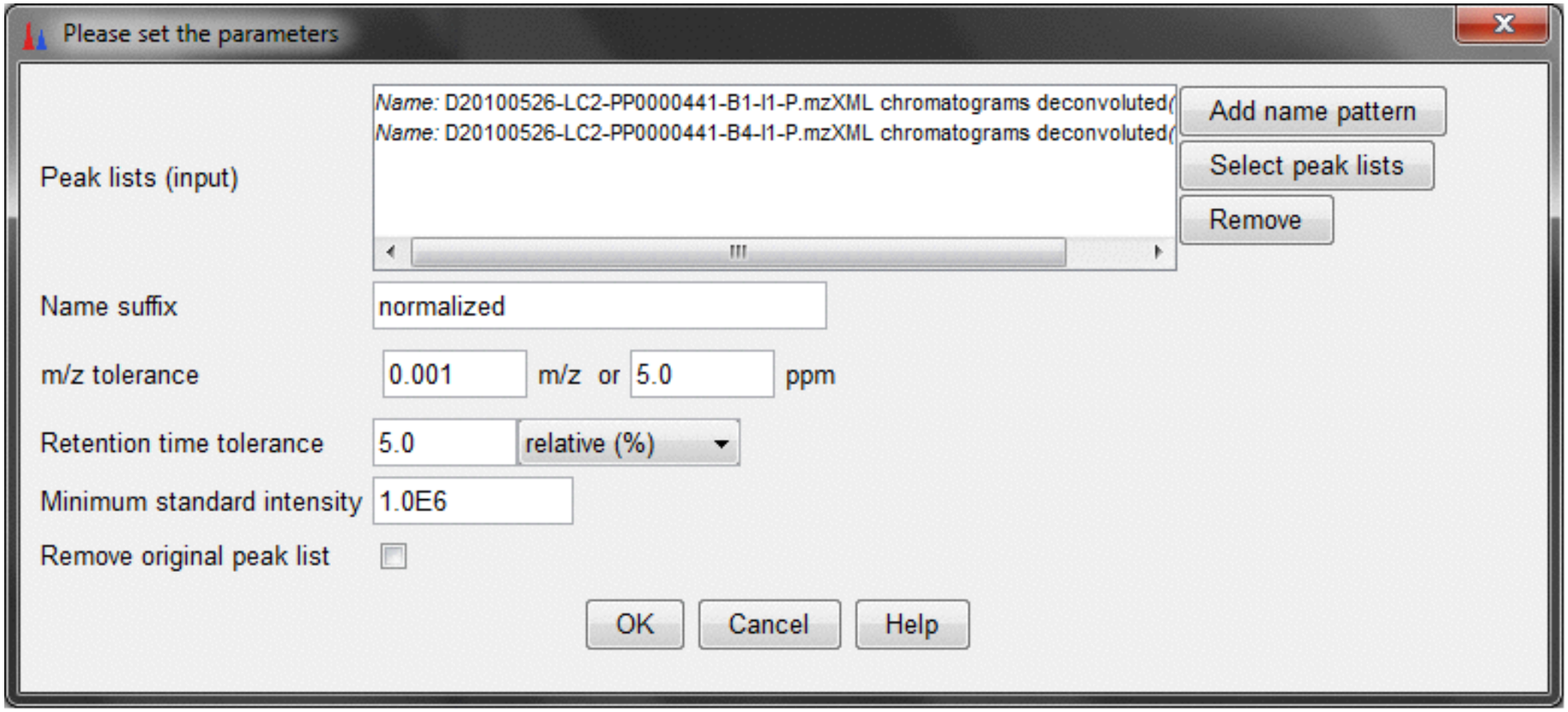
Alignment

Retention time normalization



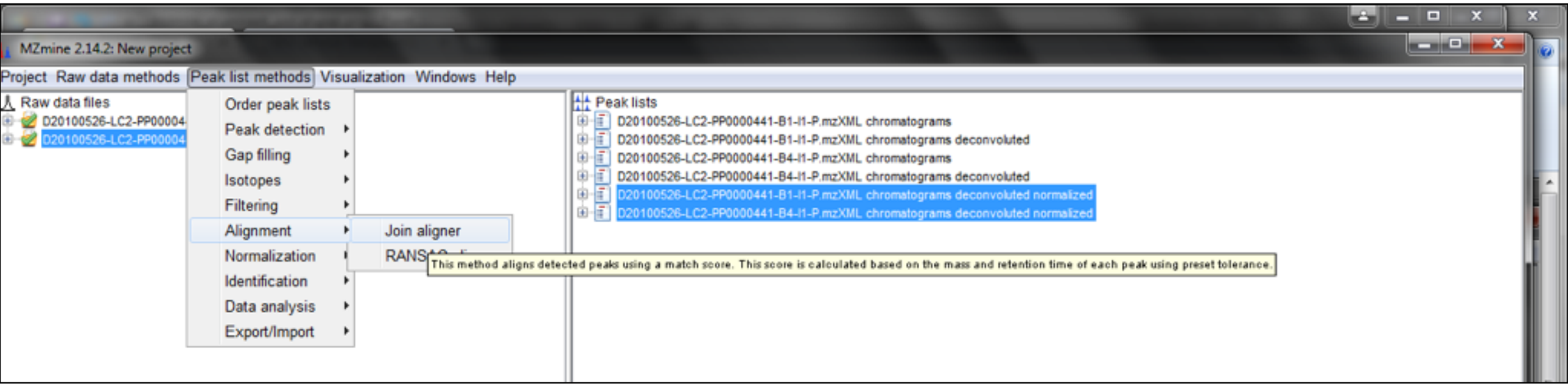
Alignment

Retention time normalization: set parameters



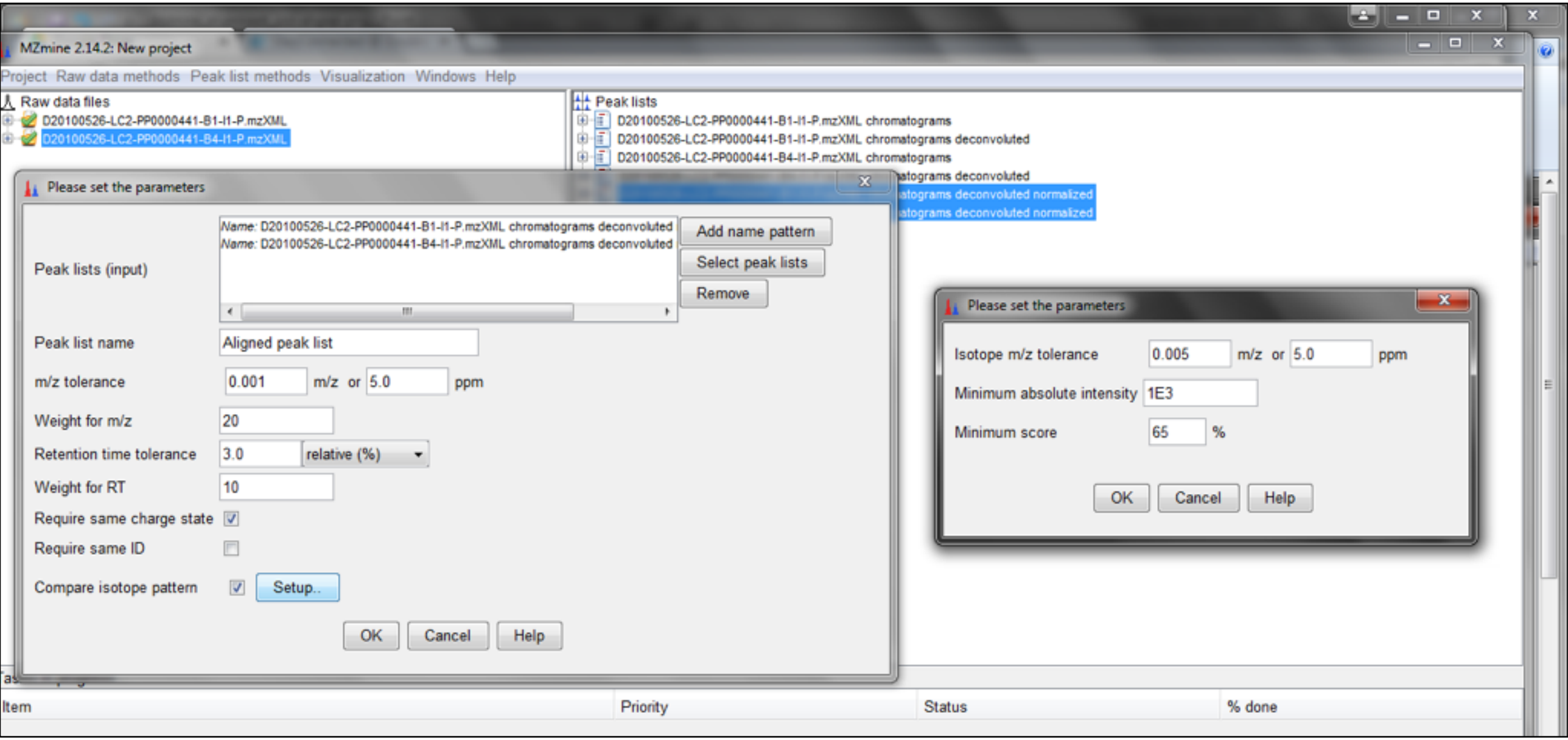
Alignment

Join aligner



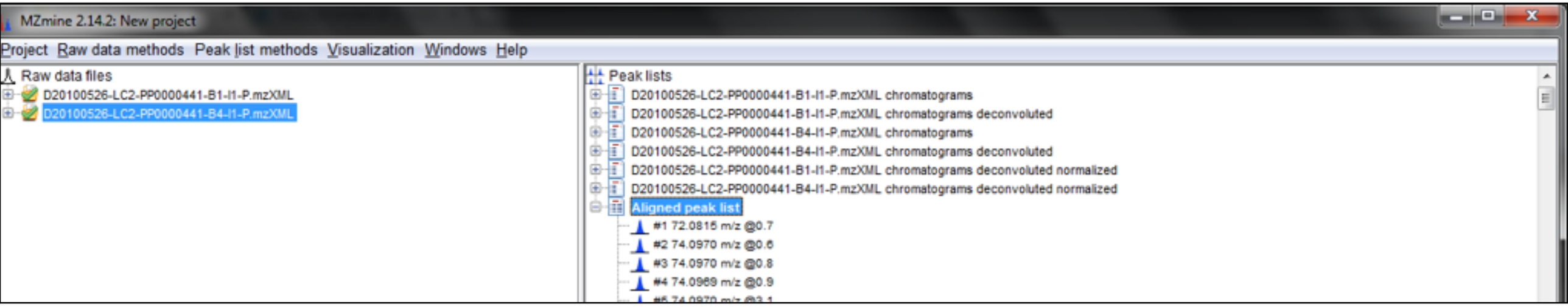
Alignment

Join aligner: set parameters



Alignment

Get aligned peak list



Alignment

Information on aligned peak list

Peak list: Aligned peak list

Windows

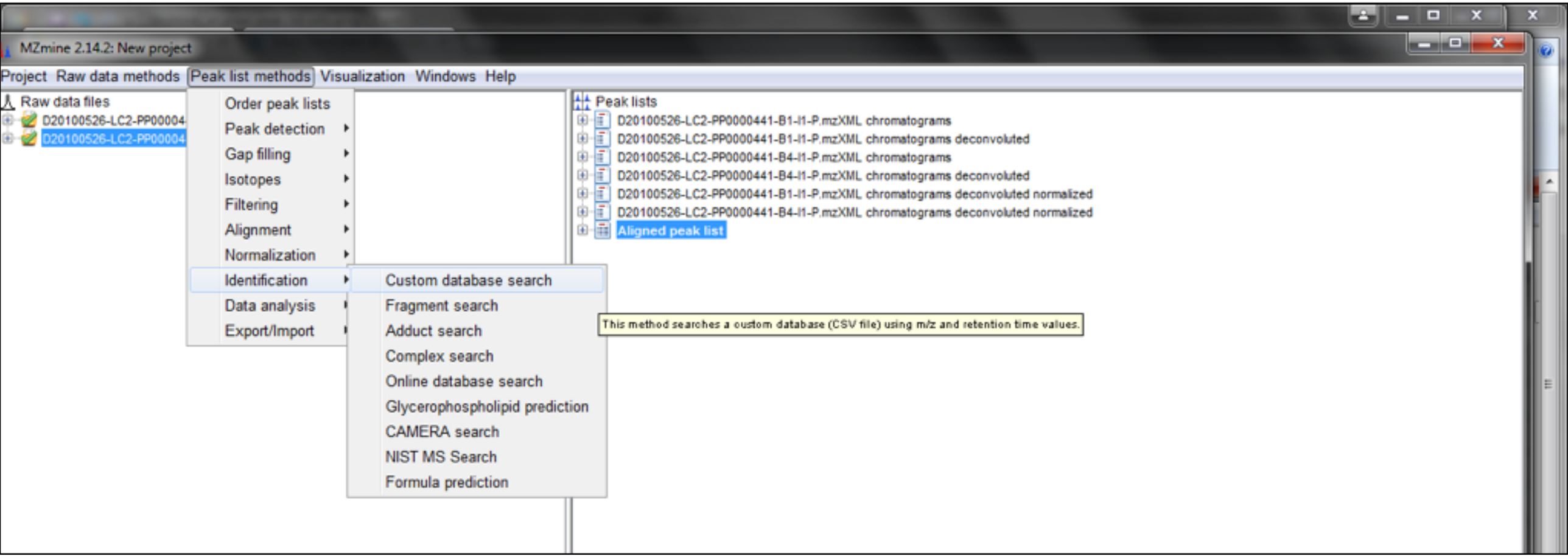
ID	Average		Identity	Comment	Peak shape	D20100526-LC2-PP0000441-B1-I1-P.mz...			D20100526-LC
	m/z	RT				Status	Height	Area	Status
1	72.0815	0.7					8.0E5	3.7E6	
2	74.0970	0.6					1.4E5	8.9E5	
3	74.0970	0.8					8.2E4	5.4E5	
4	74.0969	0.9					5.1E4	3.4E4	
5	74.0970	3.1					1.6E5	5.4E5	
6	74.0971	3.4					3.6E5	1.6E6	
7	76.0948	0.7					1.9E5	8.4E5	
8	80.9485	0.5					1.3E5	6.4E5	
9	84.9603	7.8					2.3E5	1.2E6	
10	86.0972	1.5					6.4E5	6.8E6	
11	88.1124	2.8					5.1E4	7.1E4	
12	88.1127	2.9					5.2E4	7.2E4	
13	88.1127	2.9					5.5E4	2.9E5	
14	88.1128	3.5					9.5E4	2.7E6	
15	88.1127	3.7					5.4E4	2.2E5	

Green: present

Red: absent

Peak identification

Options



Peak identification

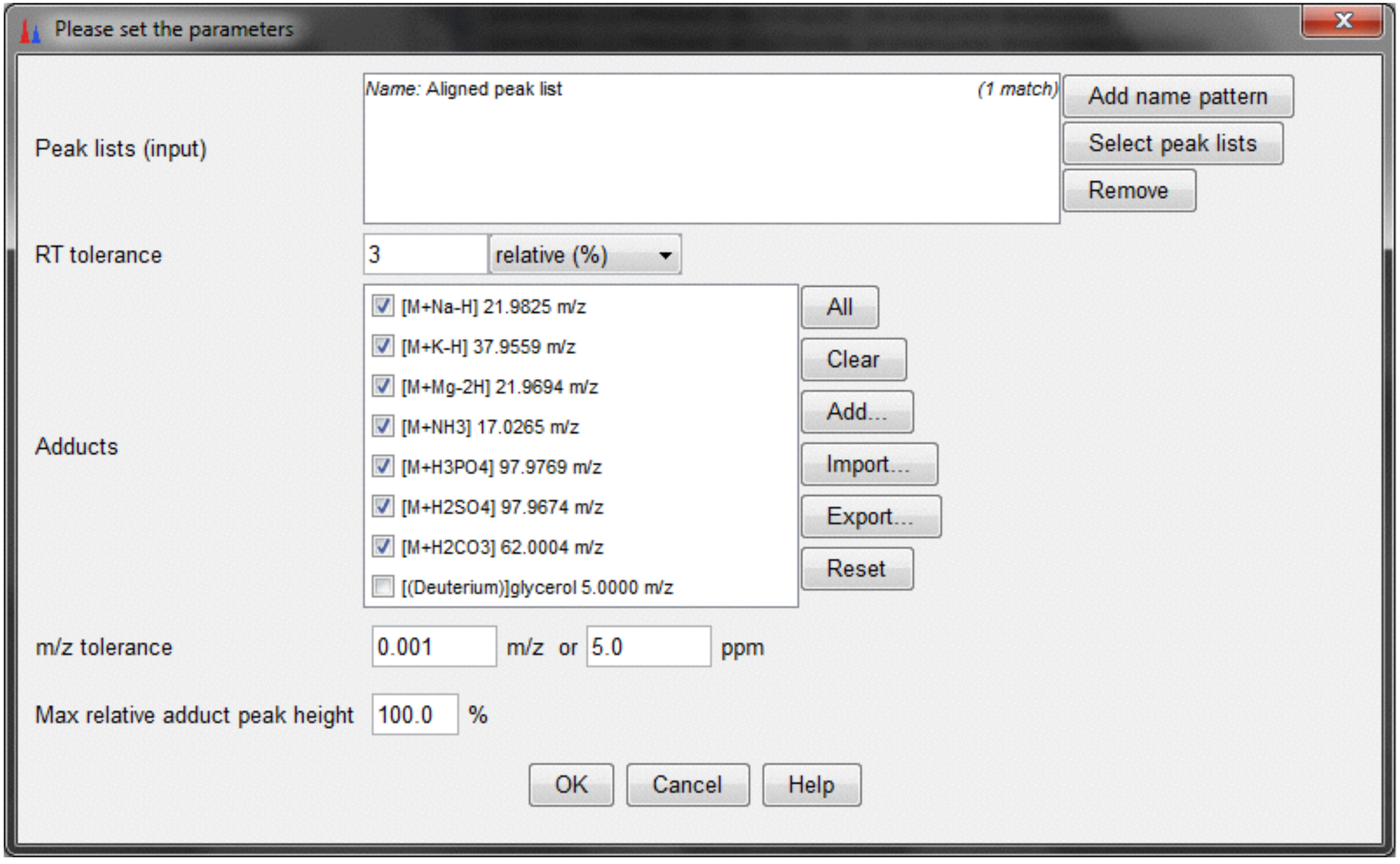
By searching custom databases

1	ID	m/z	Retention	Identity	Formula
28	C00041	90.05496	1.45	alanine [M+H]	C3H7NO2
29	na	161.0921	1.71	alanine-alanine [M+H]	C6H12N2O3
30	C01551	159.0513	1.62	allantoin [M+H]	C4H6N4O3
31	C06464	181.0707	1.52	altrose [M+H]	C6H12O6
32	C00216 C00259	151.0601	1.62	arabinose [M+H]	C5H10O5
33	C01112	231.0264	1.56	arabinose 5 phosphate [M+H]	C5H11O8P
34	C00532	153.0758	1.51	arabitol [M+H]	C5H12O5
35	C00792	175.119	1.37	arginine [M+H]	C6H14N4O2
36	C00049	134.0448	1.45	aspartic acid [M+H]	C4H7NO4
37	C00099	90.05496	1.38	beta-alanine [M+H]	C3H7NO2
38	C02512	115.0502	1.48	beta-cyano-l-alanine [M+H]	C4H6N2O2
39	C00719	118.0863	1.57	trimethylglycine [M+H]	C5H11NO2
40	C00308	177.0982	1.34	canavanine [M+H]	C5H12N4O3
41	C09773	363.1286	5.98	catalpol [M+H]	C15H22O10
42	C00185	343.1235	1.77	cellobiose [M+H]	C12H22O11
43	C01484	209.0961	10.94	chalcone [M+H]	C15H12O
44	C00852	355.1024	9.1	chlorogenic acid [M+H]	C16H18O9

an example custom database

Peak identification

By adduct search: set parameters



Peak identification

Adduct search results

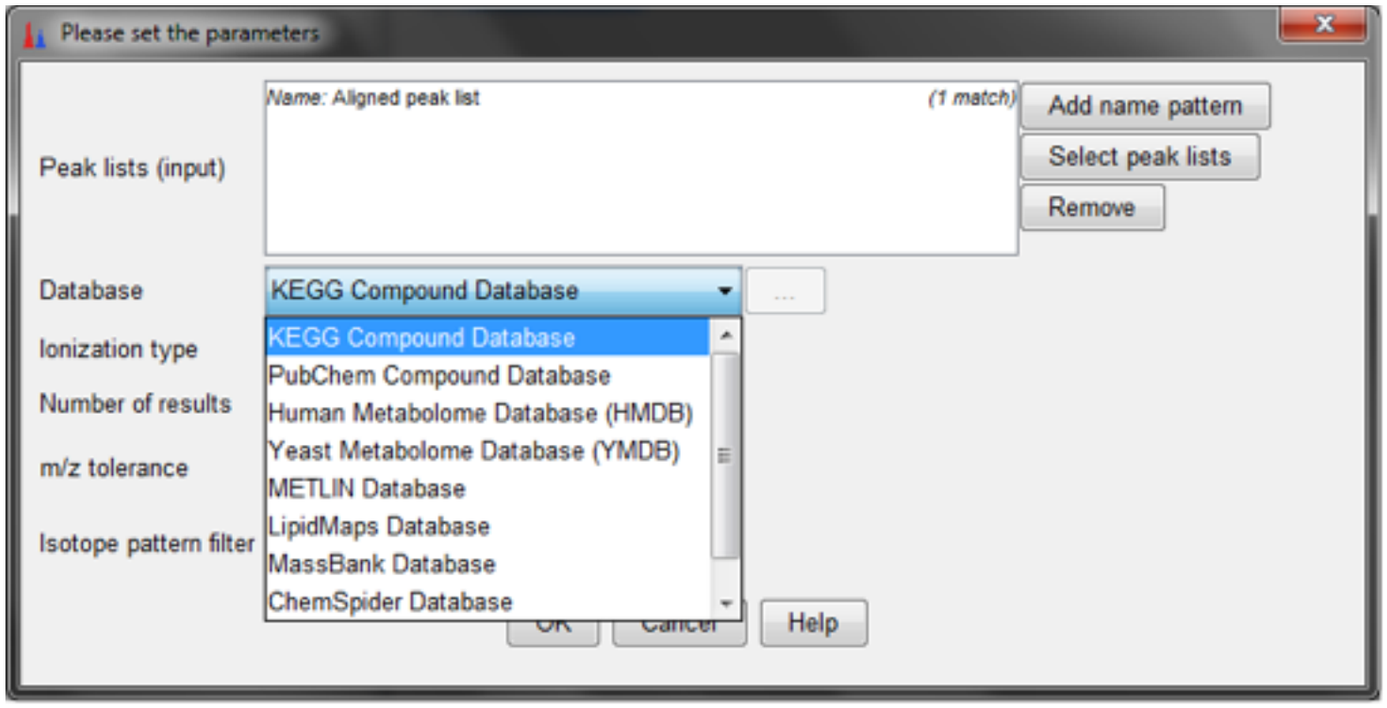
Peak list: Aligned peak list

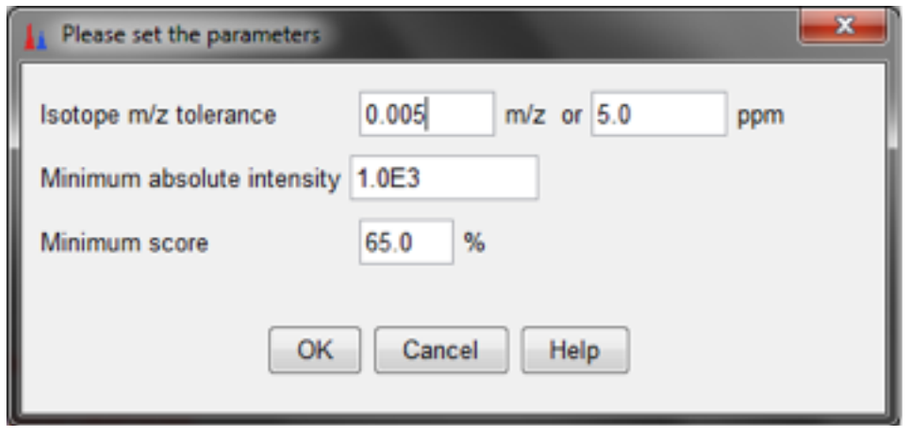
Windows

ID	Average		Identity	Comment	Peak shape	D20100526-LC2-PP0000441-B1-I1-P.mz...			D20100526-LC2-PP0000441-B4-I1-P.mz...		
	m/z	RT				Status	Height	Area	Status	Height	Area
41	121.0879	0.6					1.7E5	4.5E5			
42	123.1037	0.7					6.1E5	2.7E6		7.9E5	3.8E6
43	128.1082	2.6					4.0E5	1.5E6		5.4E5	1.6E6
44	130.0865	0.8					1.2E5	3.8E5		1.6E5	5.2E5
45	130.0506	1.0					2.2E5	1.6E6		3.1E5	1.7E6
46	132.1024	1.5					2.5E6	2.8E7			
47	132.2032	1.5					1.8E5	1.7E6			
48	133.1057	1.5					1.9E5	6.2E5		1.6E5	2.2E5
49	133.0617	3.0					5.4E5	1.3E6		9.2E5	2.4E6
50	134.0212	7.8					1.9E5	8.8E5		7.9E5	3.3E6
51	136.0486	0.6	[M+Na-H] 21.9825 m/z adduct of 114.0669 m/z				1.7E5	3.3E5		2.2E5	4.9E5
52	137.1190	1.5					1.2E5	1.1E6			
53	138.0530	0.6	[M+Na-H] 21.9825 m/z adduct of 116.0713 m/z				7.6E4	1.8E5		2.1E5	5.7E5
54	138.0492	0.9					5.6E4	3.7E4		6.8E4	1.2E5
55	138.1228	1.5					1.8E6	2.0E7			
56	138.2275	1.4					6.1E4	2.0E5			
57	138.2274	1.5					1.2E5	6.9E5			
58	140.0689	0.5					1.3E5	2.8E5		1.7E5	3.3E5
59	140.0686	0.7	[M+Na-H] 21.9825 m/z adduct of 118.0866 m/z				2.5E5	1.0E6		3.5E5	1.4E6
60	144.1027	6					3.0E5	8.4E5		5.7E5	1.5E6
61	145.0855	0.7	[M+Na-H] 21.9825 m/z adduct of 123.1037 m/z				6.3E4	1.6E5			

Peak identification

By online database search





Peak identification

Online database search results

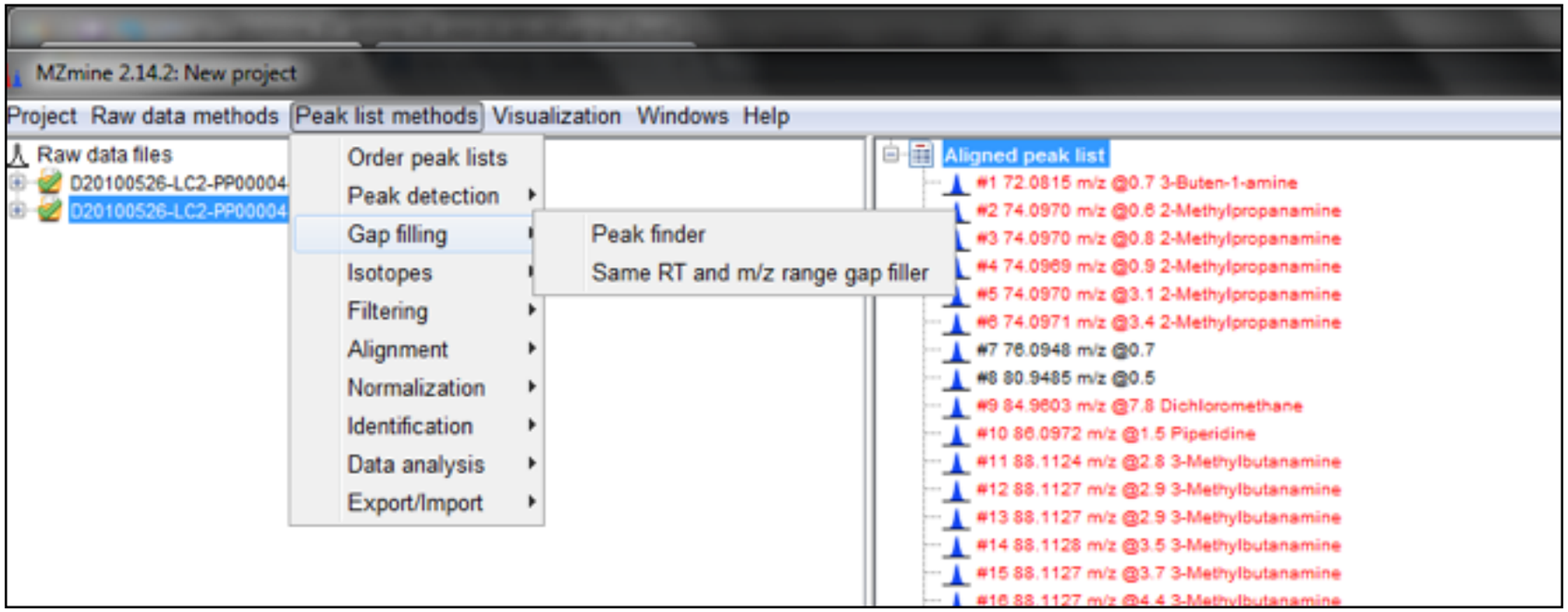
Peak list: Aligned peak list

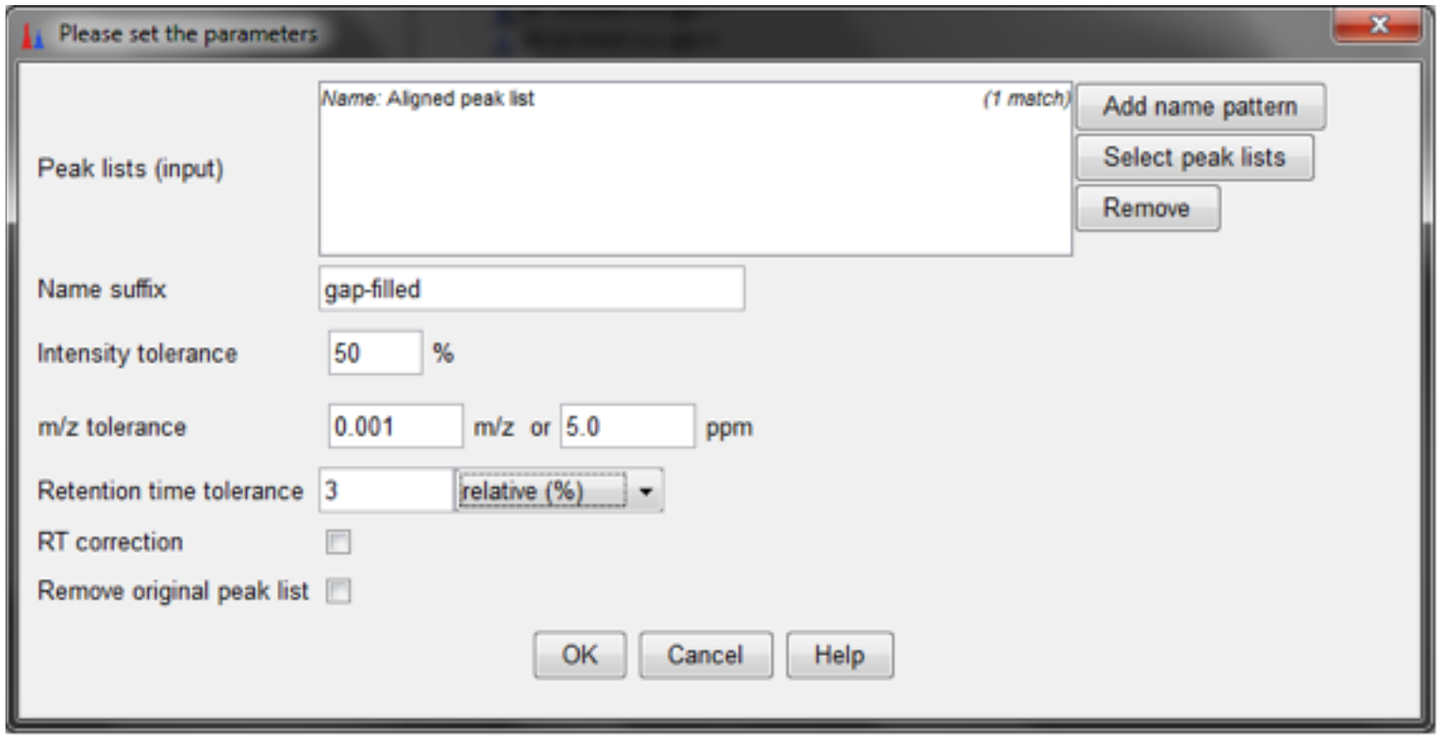
Windows

ID	Average		Identity	Comment	Peak shape	D20100526-LC2-PP0000441-B1-I1-P.mz...			D20100526-LC2-PP0000441-B4-I1-P.mz...		
	m/z	RT				Status	Height	Area	Status	Height	Area
43	128.1082	2.6					4.0E5	1.5E6		5.4E5	1.6E6
44	130.0865	0.8	L-Pipecolate				1.2E5	3.8E5		1.6E5	5.2E5
45	130.0506	1.0	4-Oxoproline				2.2E5	1.6E6		3.1E5	1.7E6
46	132.1024	1.5	L-Leucine				2.5E6	2.8E7			
47	132.2032	1.5					1.8E5	1.7E6			
48	133.1057	1.5					1.9E5	6.2E5		1.6E5	2.2E5
49	133.0617	3.0	L-Asparagine				5.4E5	1.3E6		9.2E5	2.4E6
50	134.0212	7.8					1.9E5	8.8E5		7.9E5	3.3E6
51	136.0486	0.6	[M+Na-H] 21.9825 m/z adduct of 114.0669 m/z				1.7E5	3.3E5		2.2E5	4.9E5
52	137.1191	1.5					1.2E5	1.1E6			
53	138.0534	0.6	[M+Na-H] 21.9825 m/z adduct of 116.0713 m/z				7.6E4	1.8E5		2.1E5	5.7E5
54	138.0492	0.9					5.6E4	3.7E4		6.8E4	1.2E5
55	138.1228	1.5					1.8E6	2.0E7			
56	138.2275	1.4					6.1E4	2.0E5			
57	138.2274	1.5					1.2E5	6.9E5			
58	140.0689	0.5					1.3E5	2.8E5		1.7E5	3.3E5
59	140.0686	1.7	[M+Na-H] 21.9825 m/z adduct of 118.0866 m/z				2.5E5	1.0E6		3.5E5	1.4E6
60	144.1027	0.6	(E)-4-(Trimethylammonio)but-2-enoate				3.0E5	8.4E5		5.7E5	1.5E6

Gap filling

Options and parameters





Gap filling

Results and visualization options

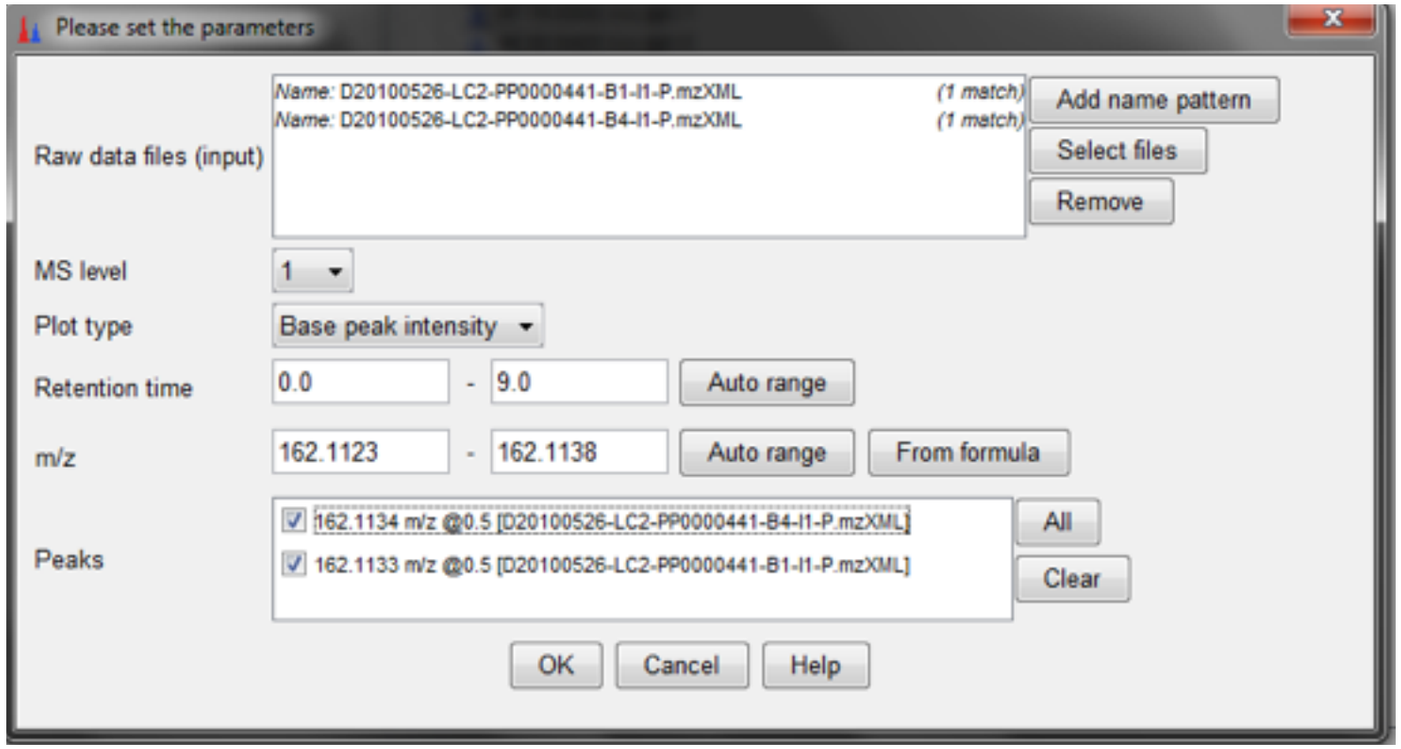
Peak list: Aligned peak list

Windows

ID	Average		Identity	Comment	Peak shape	D20100526-LC2-PP0000441-B1-I1-P.mz...			D20100526-LC2-PP0000441-B4-I1-P.mz...		
	m/z	RT				Status	Height	Area	Status	Height	Area
64	150.0589	0.9	L-Methionine				8.3E5	2.6E6		1.1E6	3.4E6
65	152.0325	1.0	[M+Na-H] 21.9825 m/z adduct of 130.0506 m/z				9.6E4	6.9E5		1.9E5	6.4E5
66	154.0843	1.5	[M+Na-H] 21.9825 m/z adduct of 132.1024 m/z				7.9E4	4.4E5			
67	155.0757	0.9					4.5E5	1.5E6		6.0E5	1.6E6
68	157.0735	0.8					5.8E4	3.7E4		8.7E4	2.4E5
69	160.1043	1.5	[M+Na-H] 21.9825 m/z adduct of 138.1228 m/z				5.4E4	1.5E5			
70	162.1133	0.5	L-Carnitine				3.8E				
71	165.0549	1.3	Phenylpyruvate				1.9E				
72	165.0350	7.8					2.8E				
73	166.0870	2.6	L-Phenylalanine				3.9E				
74	166.0875	3.5					7.8E				
75	167.0905	2.6					4.4E				
76	167.0905	3.5					8.1E4	6.3E4			
77	169.0363	1.0	Urate				6.2E5	3.6E6			
78	174.1140	2.6					3.0E5	1.1E6		3.7E5	1.1E6
79	175.1170	2.6					3.1E6	1.2E7		3.8E6	1.3E7
80	181.0729	2.7	Theophylline				1.0E5	1.7E5			
81	182.0823	1.3					1.5E6	7.9E6			
82	183.0851	1.3					1.6E5	7.2E5			
83	183.1630	0.6					9.0E4	5.0E5			

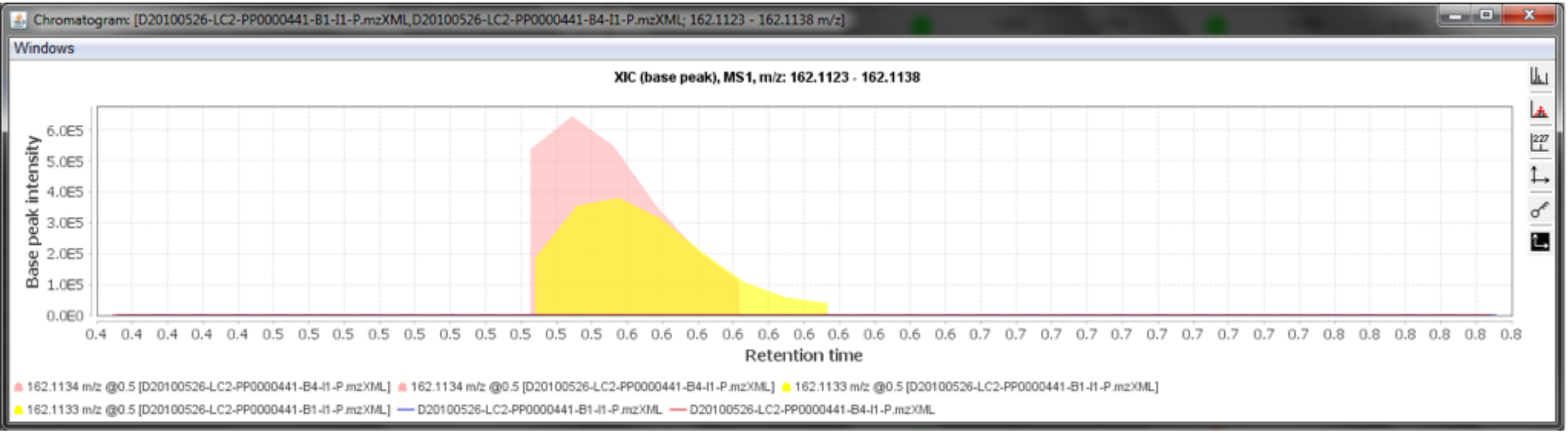
Gap filling

Set visualization parameters



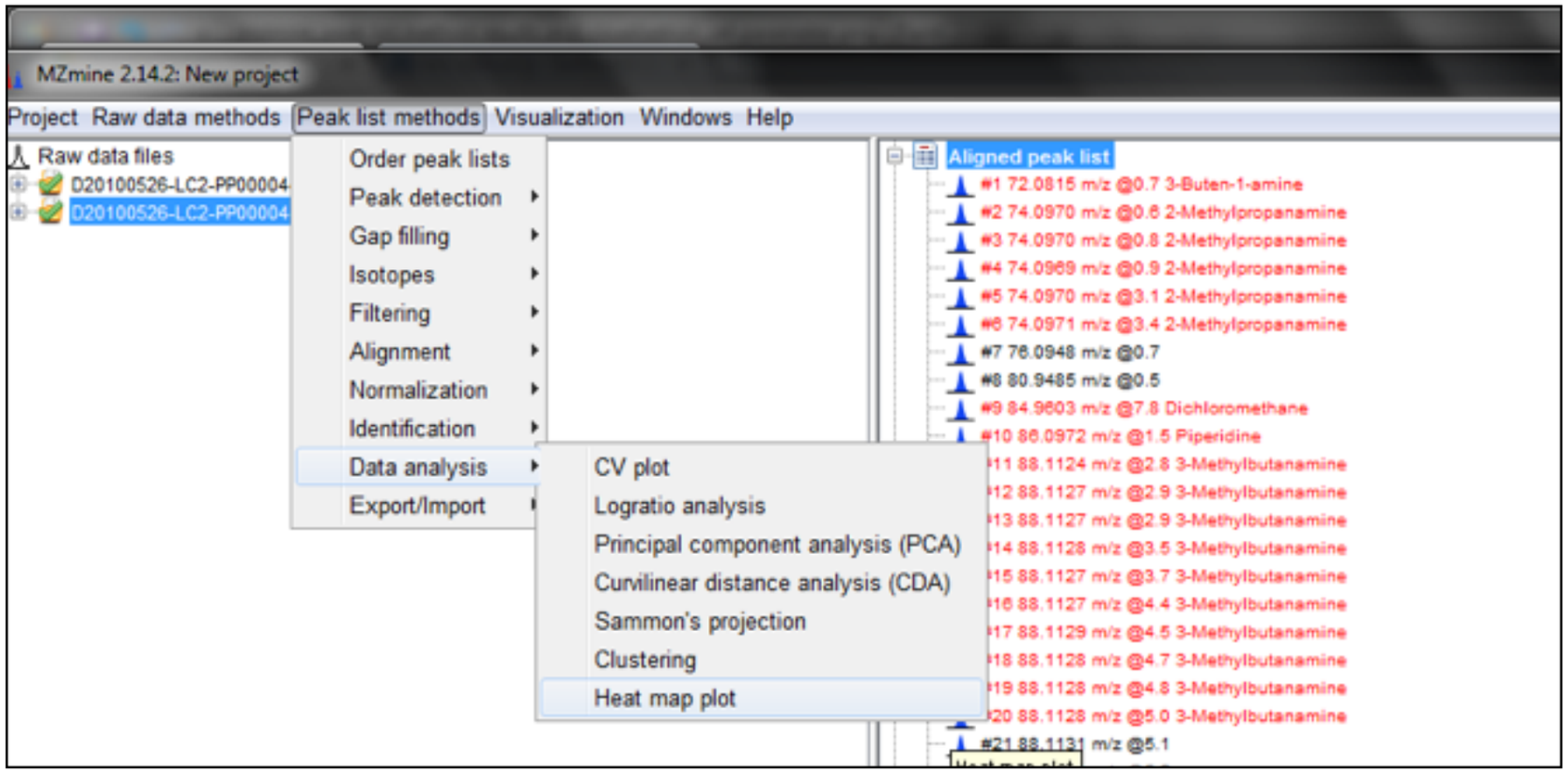
Gap filling

Visualization



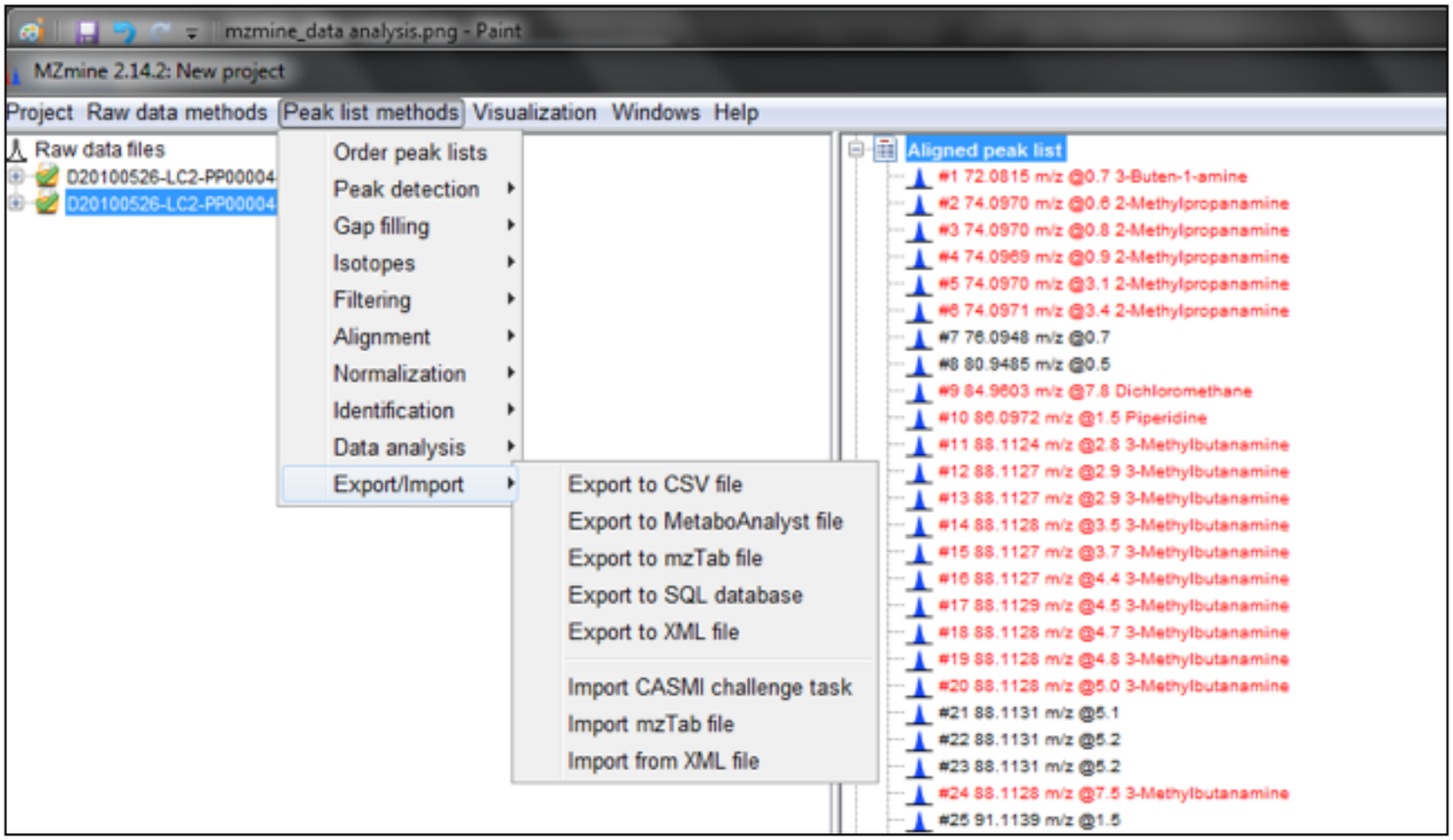
Data analysis

Options



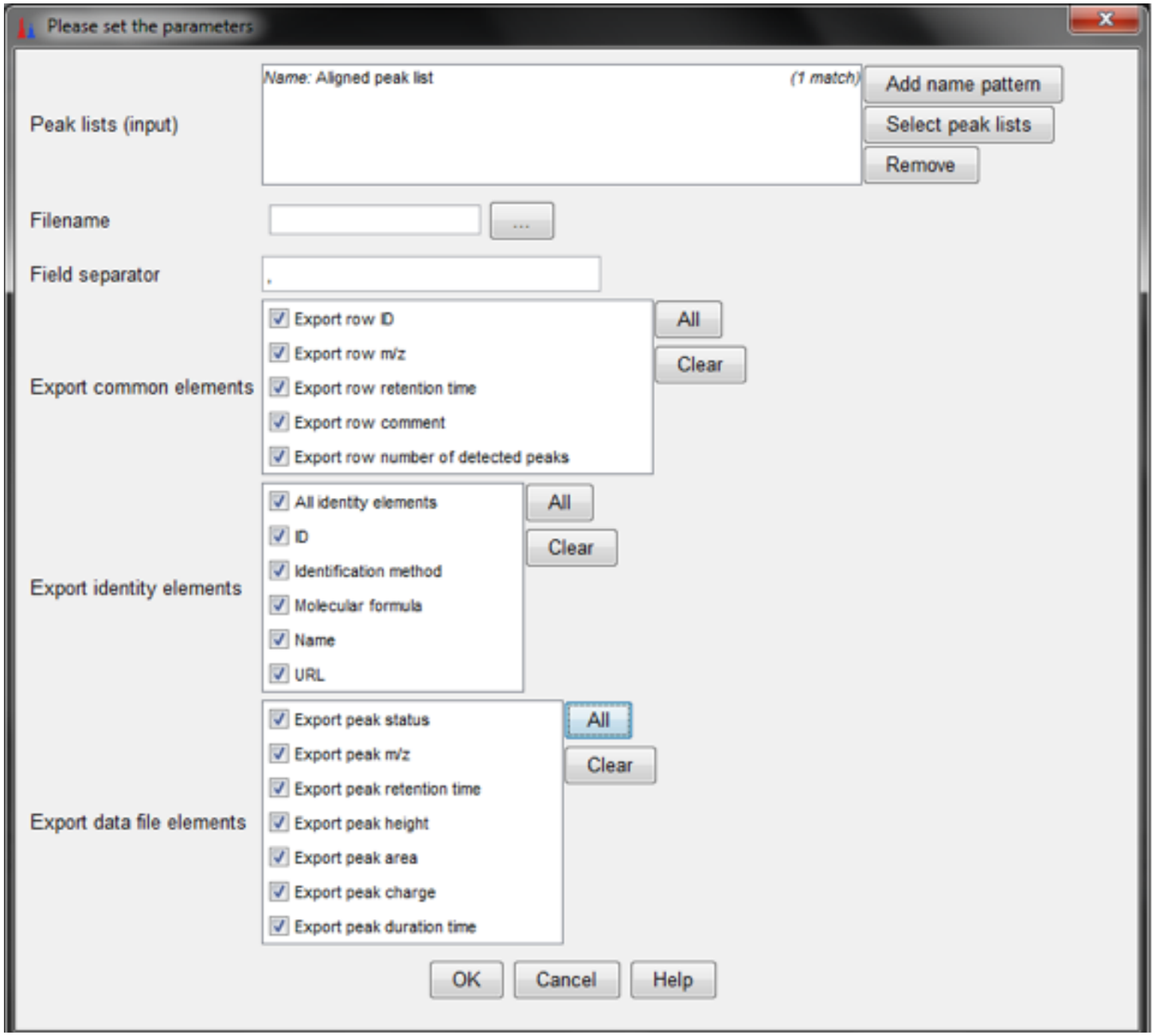
Data export

Options



Data export

Parameters



Data export

csv format

	row reten	row comn	row numb	All identit	ID	Identifica	Molecular	Name	URL	D2010052(	D20100526-LC2-PP0000441	D20100526-LC2-P	D2010052(	D2010052(	D2010052(	D2010052(	D2010052(	D2010052(	D20
2	0.724		2		C12244	KEGG sear	C4H9N	3-Buten-1	http://ww	DETECTED	72.08156586	0.723997543	800532	3677302	0	0.11805	DETECTED	72.08146	0.7
3	0.61775		1		C02787	KEGG sear	C4H11N	2-Methylp	http://ww	DETECTED	74.0969696	0.61775	135578	894695.8	0	0.141667	UNKNOW	0	
4	0.806632		1		C02787	KEGG sear	C4H11N	2-Methylp	http://ww	DETECTED	74.09701538	0.806632212	81876	540370.3	0	0.141667	UNKNOW	0	
5	0.912881		1		C02787	KEGG sear	C4H11N	2-Methylp	http://ww	DETECTED	74.09690475	0.912880769	50849	33958.16	0	0.0118	UNKNOW	0	
6	3.065793		1		C02787	KEGG sear	C4H11N	2-Methylp	http://ww	DETECTED	74.09701157	3.065792701	164087	541801.6	0	0.082633	UNKNOW	0	
7	3.44419		2		C02787	KEGG sear	C4H11N	2-Methylp	http://ww	DETECTED	74.09715652	3.494018719	362028	1621563	0	0.177083	DETECTED	74.09711	3.3
8	0.724		2							DETECTED	76.09487915	0.723997543	188653	844518.9	0	0.10625	DETECTED	76.09476	0.7
9	0.458367		2							DETECTED	80.94852448	0.464983333	132814	639803.8	0	0.11805	DETECTED	80.9485	0.
10	7.813333		2		C02271	KEGG sear	CH2Cl2	Dichlorom	http://ww	DETECTED	84.96022797	7.796366667	227123	1180269	0	0.12985	DETECTED	84.96029	
11	1.491406		1		C01746	KEGG sear	C5H11N	Piperidine	http://ww	DETECTED	86.09718323	1.49140625	635862	6786687	0	0.295133	UNKNOW	0	
12	2.837807		1		C02640	KEGG sear	C5H13N	3-Methylb	http://ww	DETECTED	88.11236572	2.837806972	51294	70746.81	0	0.023617	UNKNOW	0	
13	2.873806		1		C02640	KEGG sear	C5H13N	3-Methylb	http://ww	DETECTED	88.11269379	2.87380561	51532	71944.87	0	0.023617	UNKNOW	0	
14	2.945803		1		C02640	KEGG sear	C5H13N	3-Methylb	http://ww	DETECTED	88.11274719	2.945802887	55102	293767.2	0	0.09445	UNKNOW	0	
15	3.494019		1		C02640	KEGG sear	C5H13N	3-Methylb	http://ww	DETECTED	88.11279297	3.494018719	94520	2704280	0	0.672917	UNKNOW	0	
16	3.725024		1		C02640	KEGG sear	C5H13N	3-Methylb	http://ww	DETECTED	88.11273193	3.725024455	54012	219970.6	0	0.070833	UNKNOW	0	
17	4.383776		1		C02640	KEGG sear	C5H13N	3-Methylb	http://ww	DETECTED	88.11273956	4.383775926	51651	71837	0	0.023617	UNKNOW	0	
18	4.52545		2		C02640	KEGG sear	C5H13N	3-Methylb	http://ww	DETECTED	88.1129303	4.525448148	64166	485431.6	0	0.141667	DETECTED	88.11287	4.5
19	4.735547		1		C02640	KEGG sear	C5H13N	3-Methylb	http://ww	DETECTED	88.11283875	4.735546894	56805	412476.8	0	0.129867	UNKNOW	0	
20	4.823239		1		C02640	KEGG sear	C5H13N	3-Methylb	http://ww	DETECTED	88.1128006	4.823239386	54623	216883.8	0	0.070833	UNKNOW	0	
21	4.998642		1		C02640	KEGG sear	C5H13N	3-Methylb	http://ww	DETECTED	88.11277771	4.998642059	58821	579818.9	0	0.177083	UNKNOW	0	
22	5.111397		1							DETECTED	88.1131134	5.11139713	52386	35898.43	0	0.0118	UNKNOW	0	
23	5.161505		1							DETECTED	88.11308289	5.161504599	52567	145161.7	0	0.047233	UNKNOW	0	

MetaboSearch

MetaboSearch

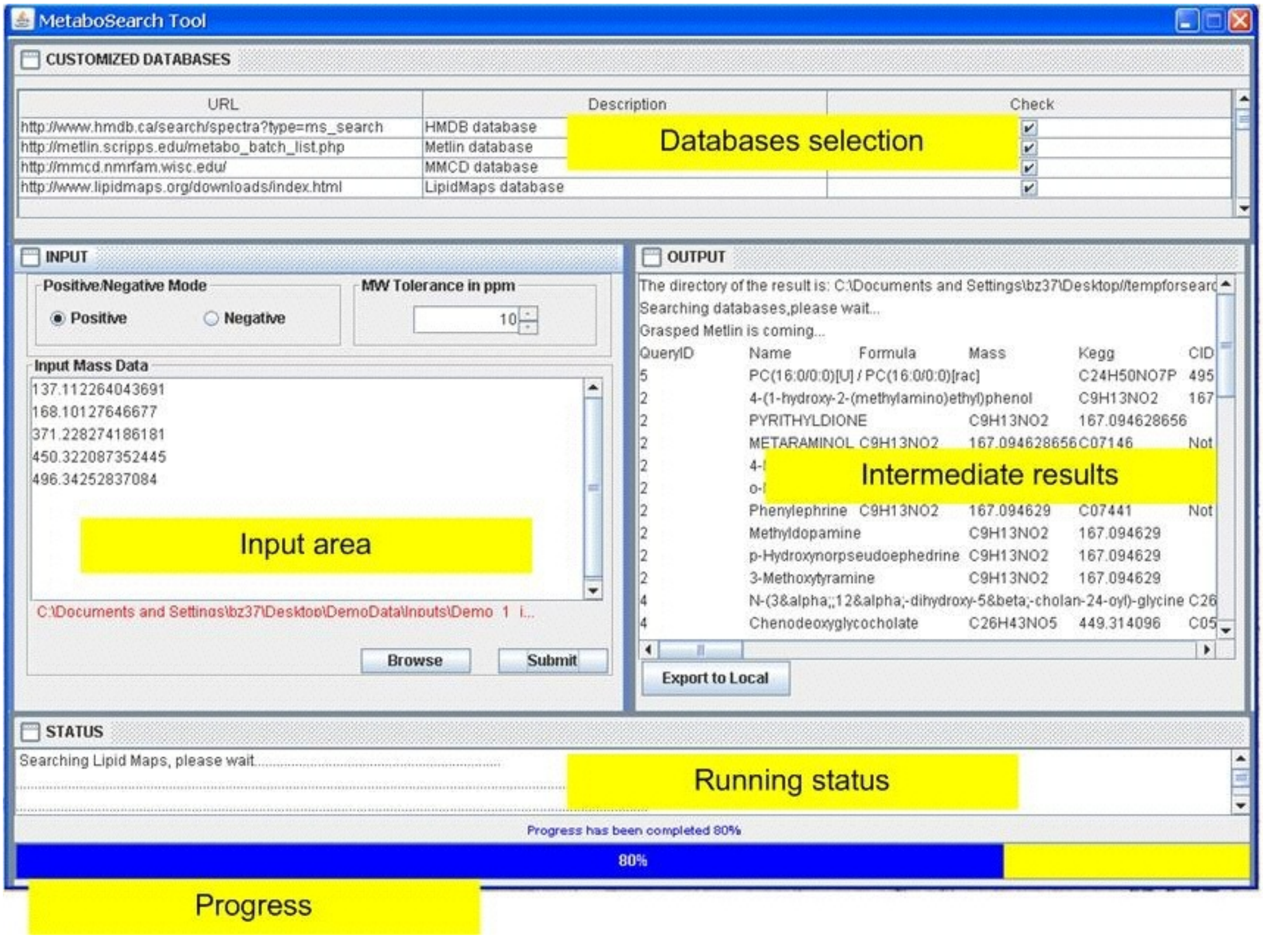
Input data format

	A	B	C
1	mz	rt	
2	62.98193	31.44898	
3	69.06956	50.27389	
4	70.06484	29.59302	
5	72.0804	34.29358	
6	73.08333	34.24128	
7	73.53049	23.18216	
8	80.94842	27.18991	
9	82.01386	23.07363	
10	82.53558	23.18671	
11	82.94562	27.18991	
12	83.01783	23.4163	

http://omics.georgetown.edu/metabosearch.html#ug

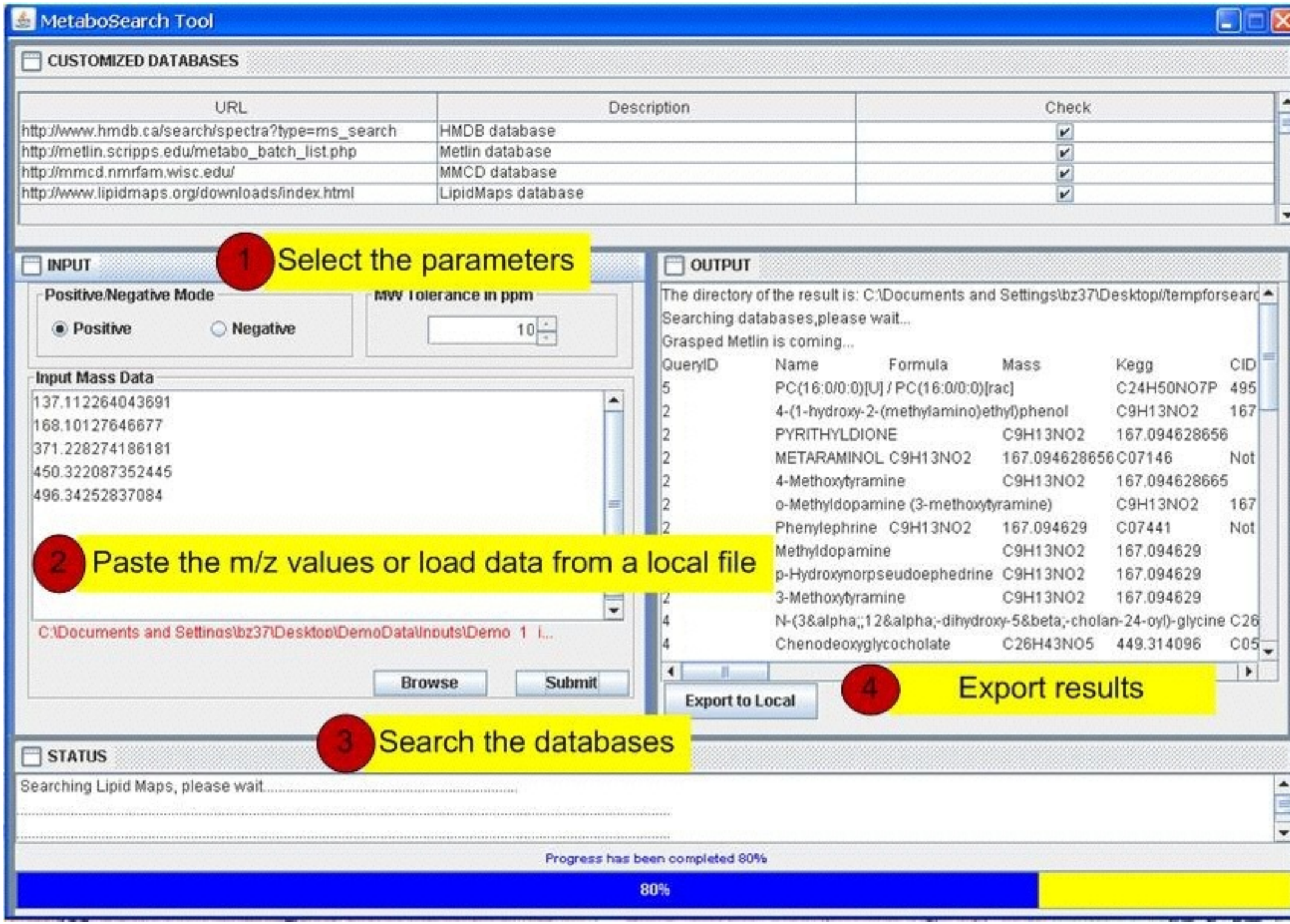
MetaboSearch

GUI



MetaboSearch

Steps



Software tools for GC-MS Metabolomics

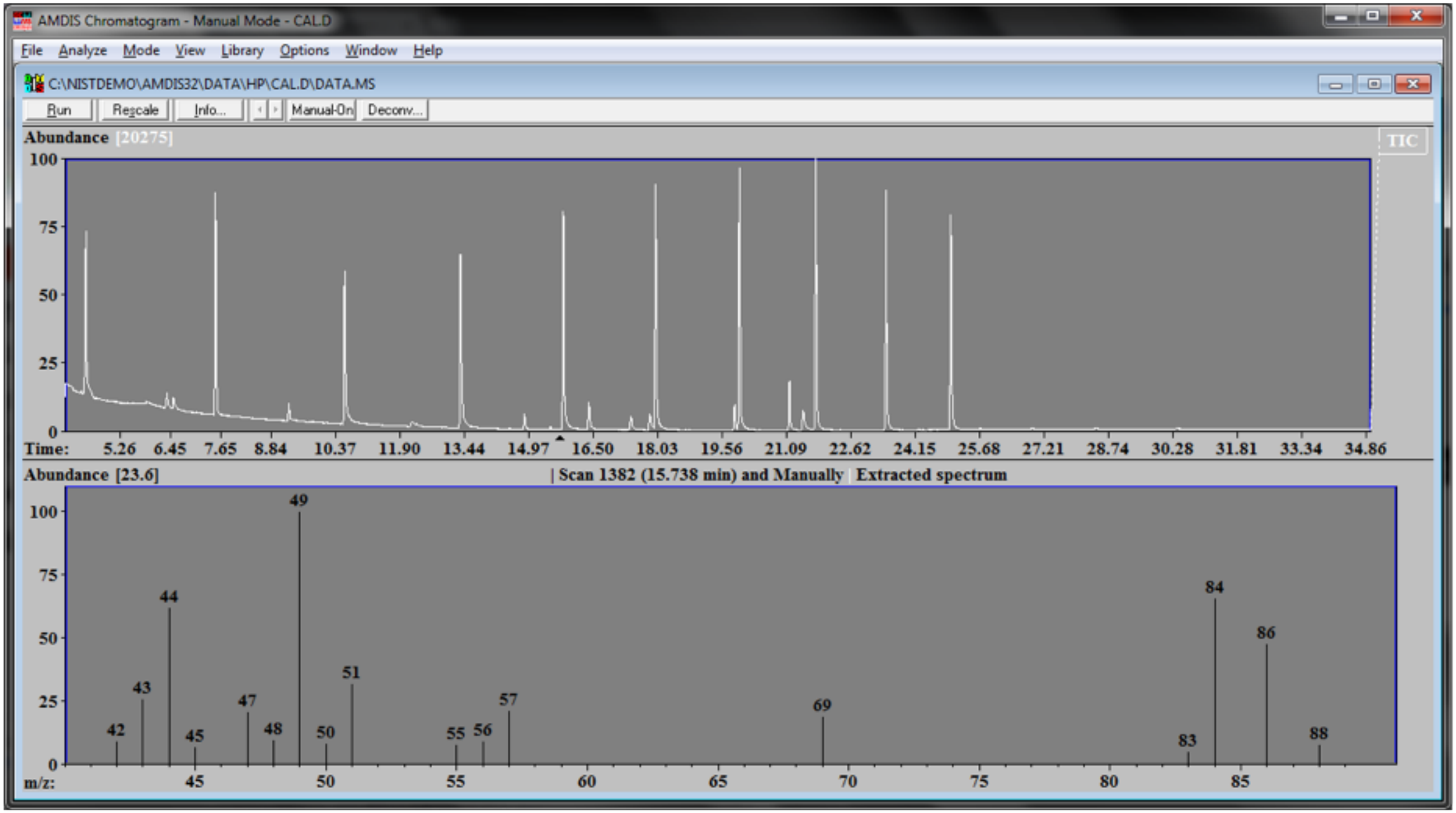
Software Tools for GC-MS

- **Commercial**
  - ChromaTOF® — LECO
  - MassHunter — Agilent

- **Free**
  - **AMDIS:** Automated Mass Spectral Deconvolution and Identification System
  - **NIST MS Search**
  - Tagfinder

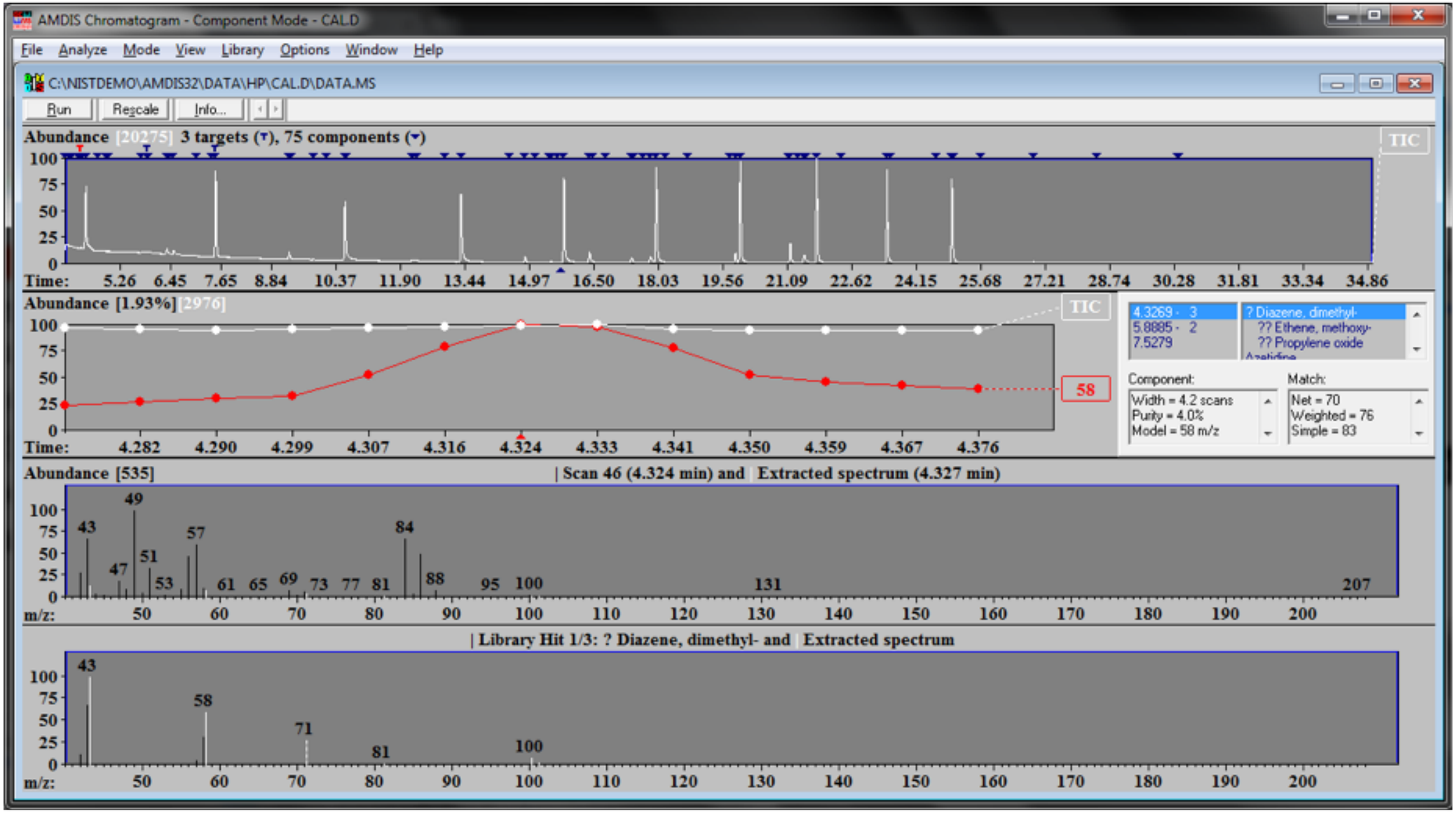
AMDIS

Raw data visualization



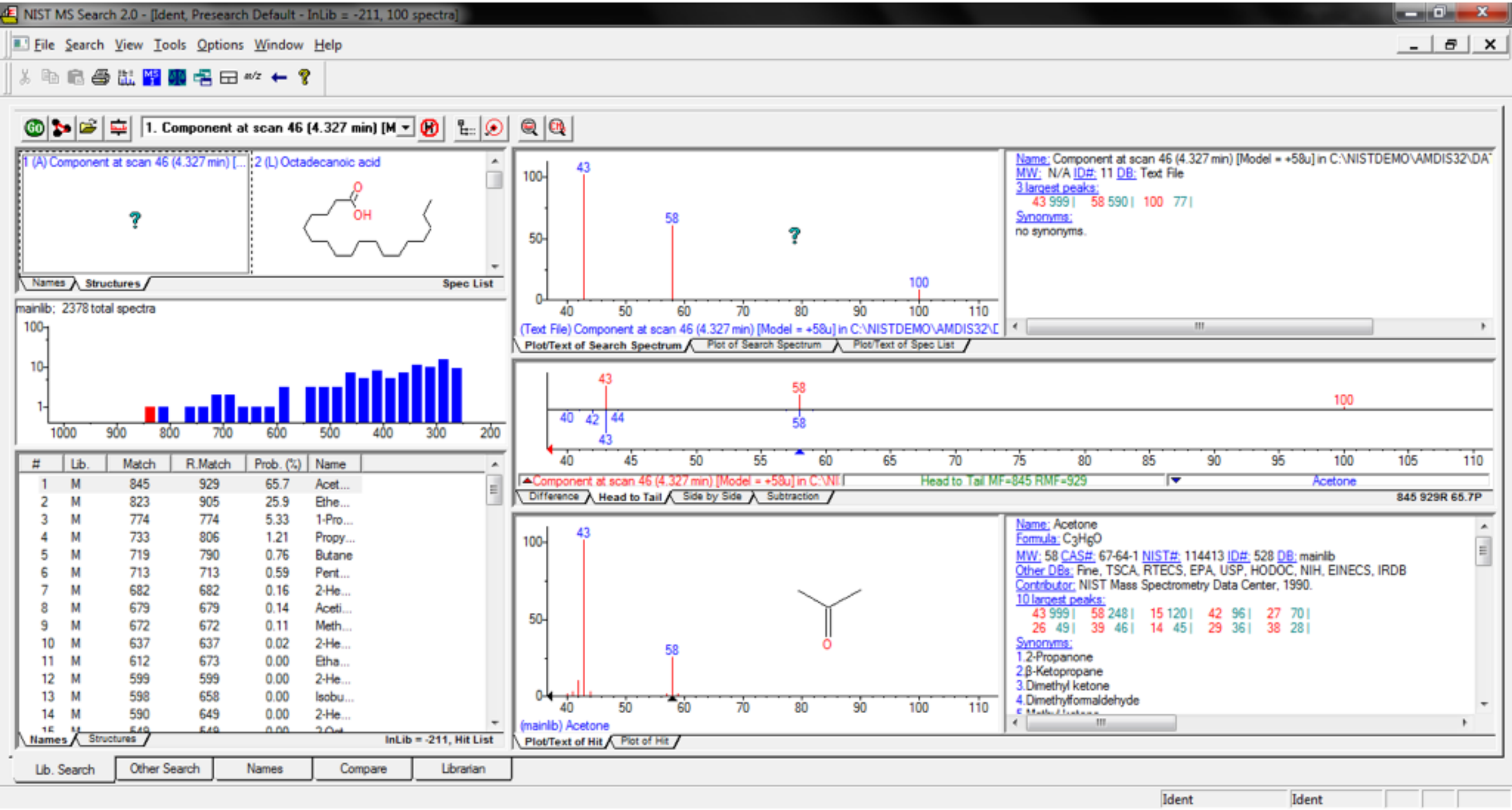
AMDIS

Result visualization



NIST MS Search

GUI



Software tools for Statistical Analysis

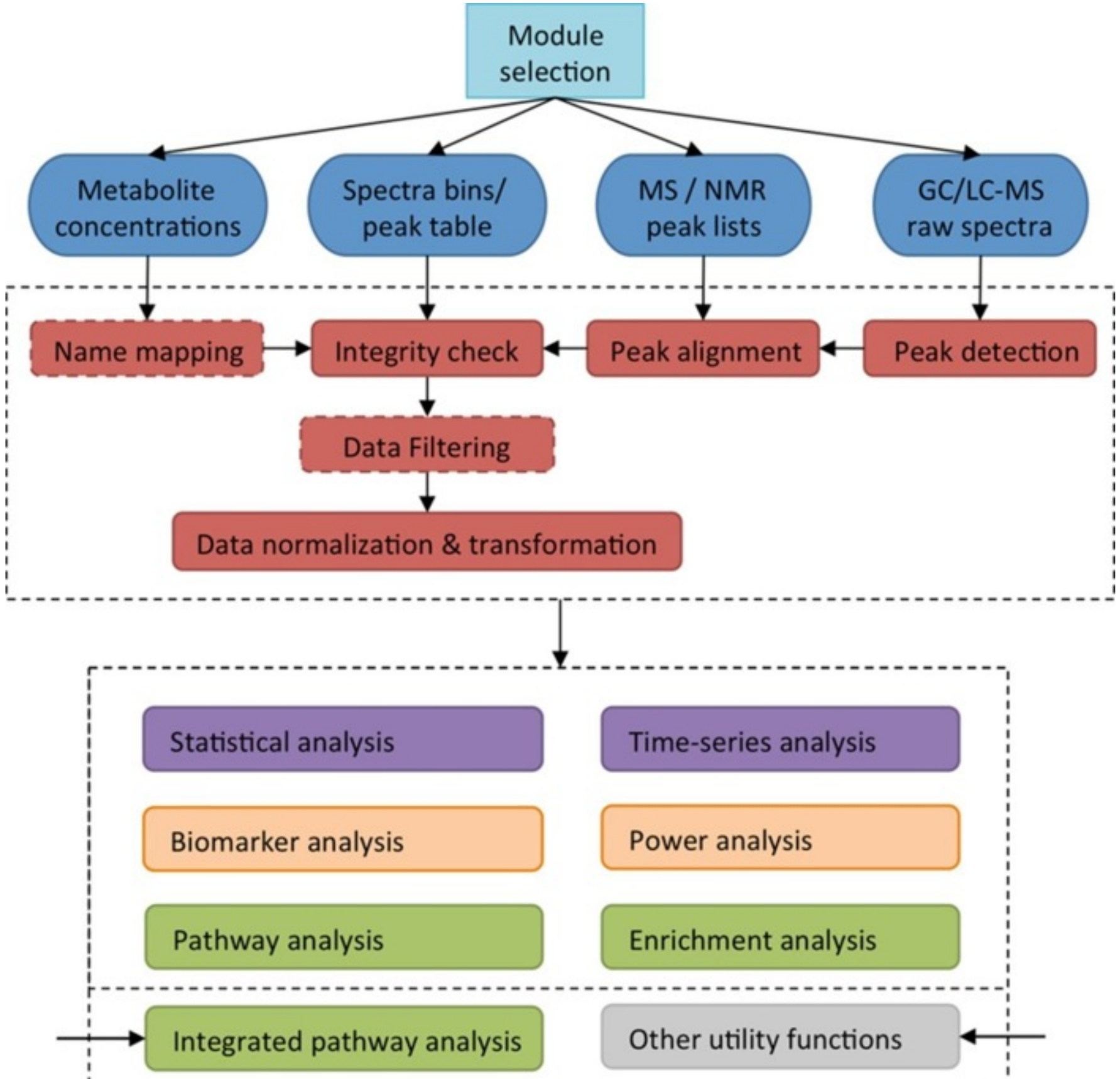
Stats and Machine Learning

- **Commercial**
  - SIMCA
  - Mass Profiler Professional (MPP) — Agilent

  - and more ……

- **Free**
  - **MetaboAnalyst**
  - R

MetaboAnalyst



Other software tools

- **COMSPARI**
  - COMparison of SPectral And Retention Information
- **MathDAMP**
  - Mathematica package for Differential Analysis of Metabolite Profiles
- **MSFACTs**
  - Metabolomics Spectral Formatting, Alignment and Conversion Tools

and many more ……

Thank You!