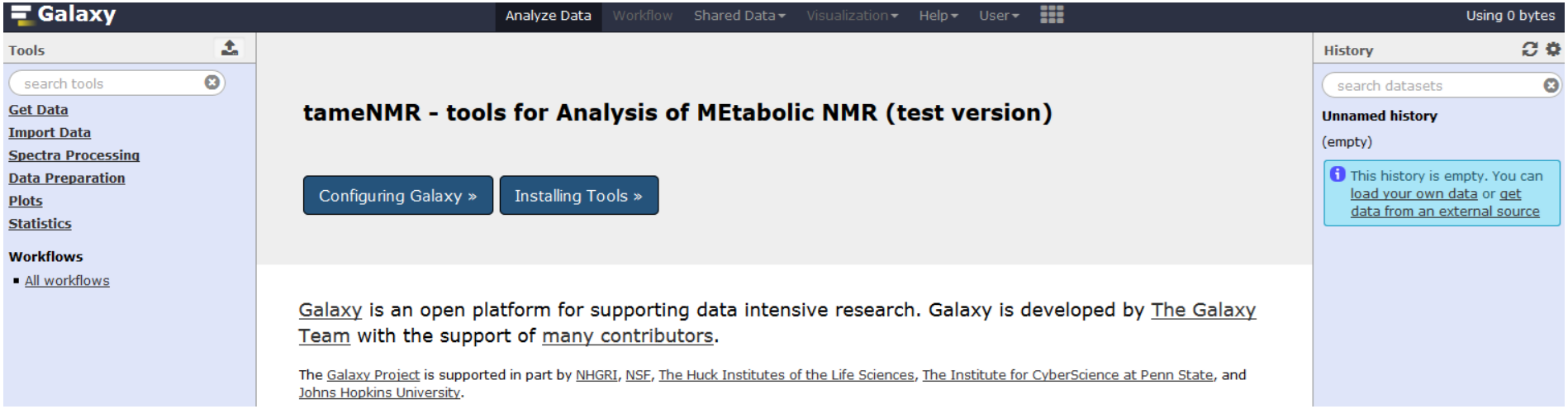
UNIVERSITY OF LIVERPOOL

@LivUniNMR

NMR Metabolomics Workshop

Appraisal of 1D spectra using tameNMR

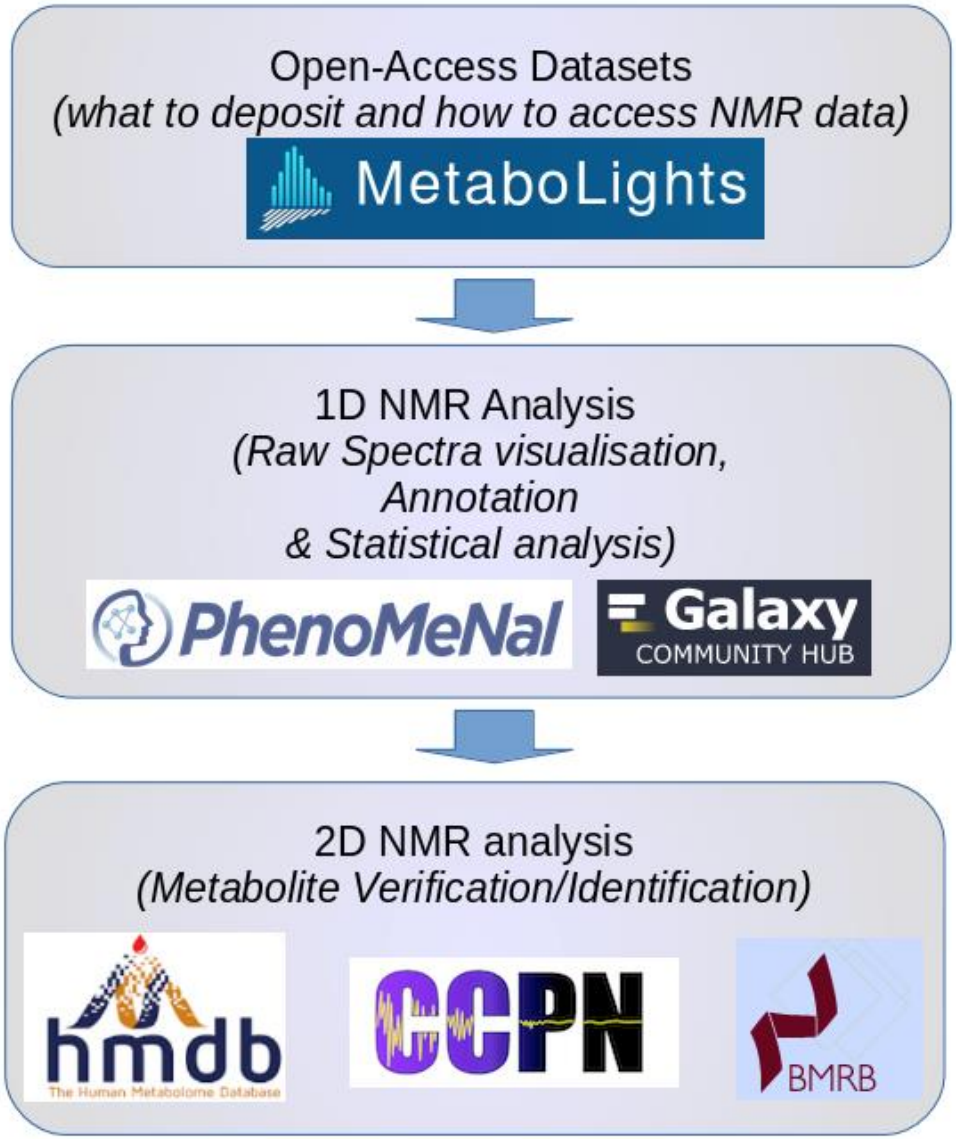


NMR Metabolomics

23rd June 2019

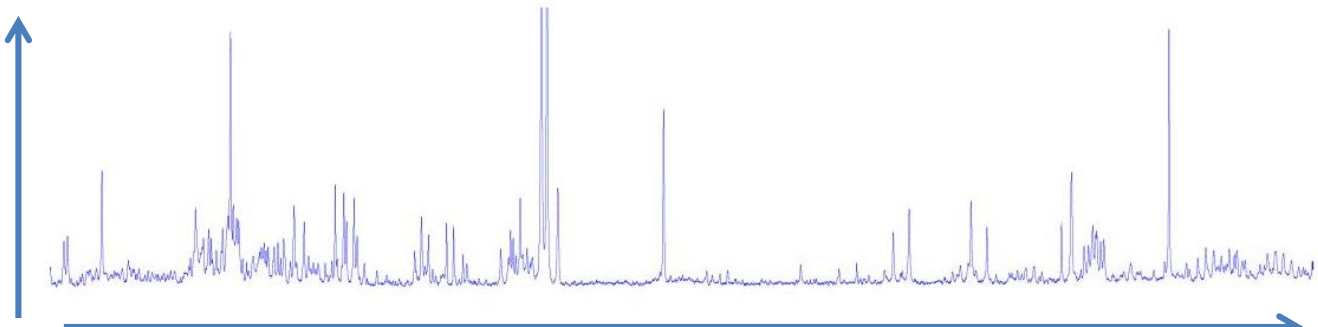
CBF
COMPUTATIONAL BIOLOGY FACILITY

Workshop Structure



Section 2. Accessing data downloaded from MetaboLights via Galaxyproject.org tool tameNMR (available via PhenoMeNal) and other tools available via workflow4metabolomics. This section will look at raw 1D 1H NMR spectra deposited and bin spectra according to annotation provided with deposition and perform simple statistics on the resultant spectra intensity files..

Spectral Formatting



Metabolites/peaks/buckets

Transformation of a Spectrum

Only useful to NMR spectroscopists using specific software

samples

intensity values

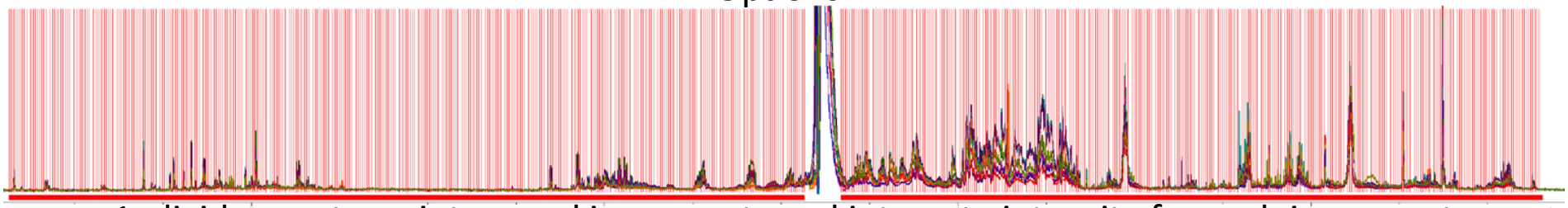
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6	Pseudo_6
2	9.99975	3.86E-05	8.23E-06	1.36E-05	3.38E-05	2E-07	2.6E-05	-5.7E-05	-3.6E-05	2.35E-05	4.48E-05	-3E-06	-4.4E-06	-5.6E-06	0.000026
3	9.99925	3.49E-05	-5.1E-07	1.4E-05	1.67E-05	1.43E-05	8.16E-06	-6.5E-05	2.02E-05	0.000026	2.89E-05	-2.1E-05	6.82E-05	-2E-06	2.15E-05
4	9.99875	-6.5E-06	6.67E-06	3.38E-05	2.69E-05	1					.48E-05	-2.1E-05	4.57E-05	2.76E-06	3.77E-05
5	9.99825	-9E-06	-5.5E-06	5.06E-05	1.35E-05	-					.83E-05	7.01E-06	1.96E-05	1.47E-05	4.65E-05
6	9.99775	-2.1E-05	-1.2E-05	5.25E-05	-3.6E-05	-					.22E-06	2.21E-05	4.1E-05	1.11E-05	4.06E-05
7	9.99725	-1.7E-05	-1.1E-05	5.91E-05	-5E-05	-2.6E-05	1.06E-05	3.35E-05	4.13E-05	3.09E-06	7.63E-06	2.58E-05	3.46E-05	2.37E-05	2.01E-05
8	9.99675	-2.5E-06	-2.1E-05	9.55E-06	-1.9E-05	-2.4E-06	1.4E-05	8.02E-06	-3.9E-07	-2.1E-05	5.63E-06	9.77E-06	2.61E-05	4.96E-05	2.65E-06
9	9.99625	5.85E-06	-1.2E-05	-2.5E-05	6.78E-06	-4.6E-06	-1.1E-06	2.29E-05	-1.5E-05	1.68E-05	1.01E-05	1.42E-05	2.21E-05	4.51E-05	1.67E-05
10	9.99575	-4.6E-07	-2.5E-05	-6.9E-06	-9E-06	-1.7E-05	1.49E-05	7.39E-05	-3.3E-05	3.42E-05	0.000017	-5.4E-06	2.76E-05	2.59E-05	6.22E-06
11	9.99525	-3E-06	-8.2E-06	-1.2E-05	-1.8E-05	-2.2E-05	1.72E-05	7.91E-05	-2.7E-05	3.5E-05	3.24E-05	-1E-05	1.79E-05	1.15E-05	7.5E-06
12	9.99475	-3.8E-06	3.99E-05	4.84E-05	-5.5E-05	-3.7E-06	2.85E-05	3.54E-05	-1.6E-06	1.9E-05	4.84E-05	-1.9E-05	6.39E-05	2.51E-05	1.82E-06

Into a Matrix of numbers.

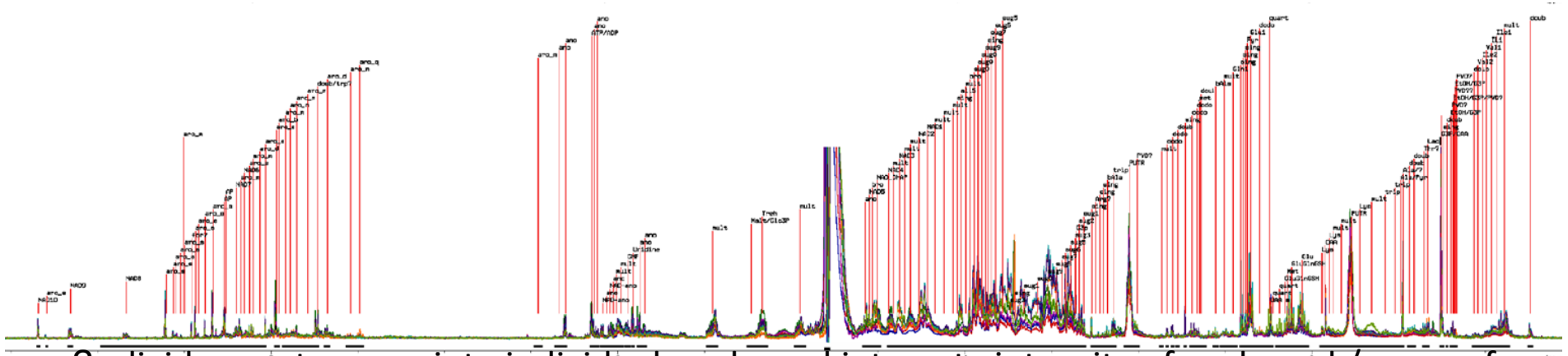
Interpretation by multiple analysts and tools

Preparing a Bucket Table/Binning Spectra

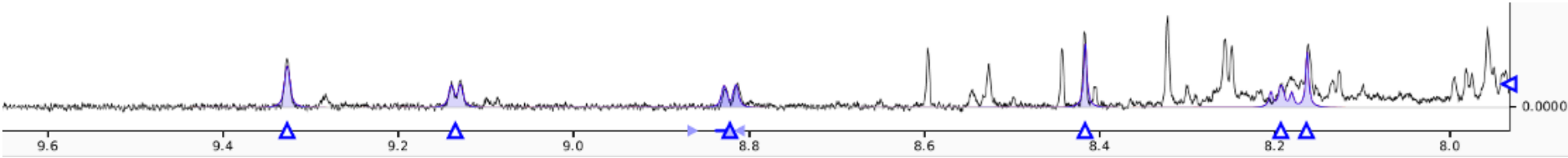
Options



- 1. divide spectrum into equal increments and integrate intensity for each increment



2. divide spectrum up into individual peaks and integrate intensity of each peak/group of peaks (requires a **pattern** file)



3. peak deconvolution via fitting wave functions (batman, Chenomx etc.)

Incremental Bucketing

Advantages:

- Quick
- Consistent
- Unbiased
- Useful for a first check

Limitations:

- No metabolite information
- Could include artefact/contaminant peaks
- Larger number of buckets

Bucketing *via* Pattern File

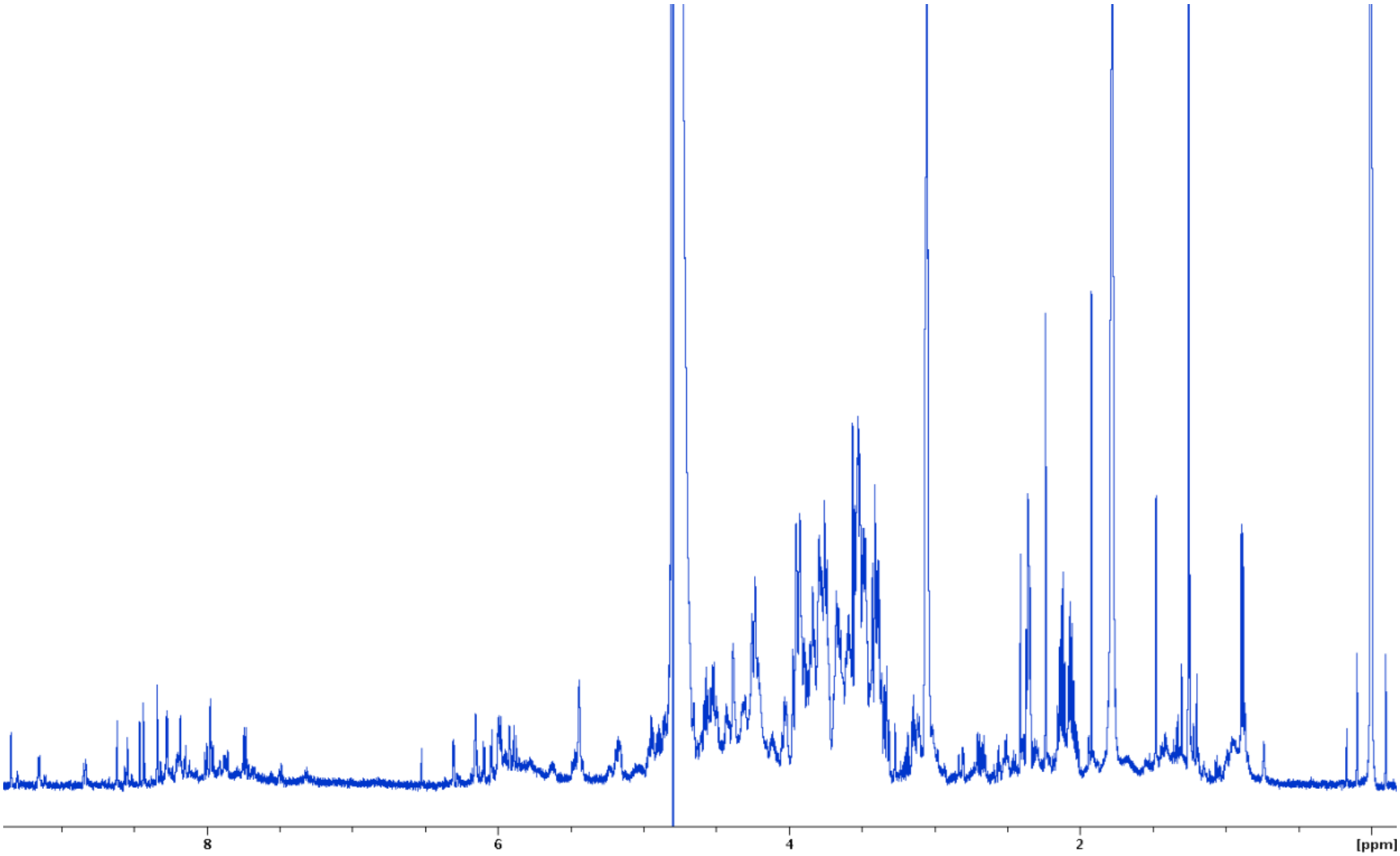
Advantages:

- All peaks represented
- Removal of noise
- Analysis can by metabolite driven
- Intensities reflect the individual molecules

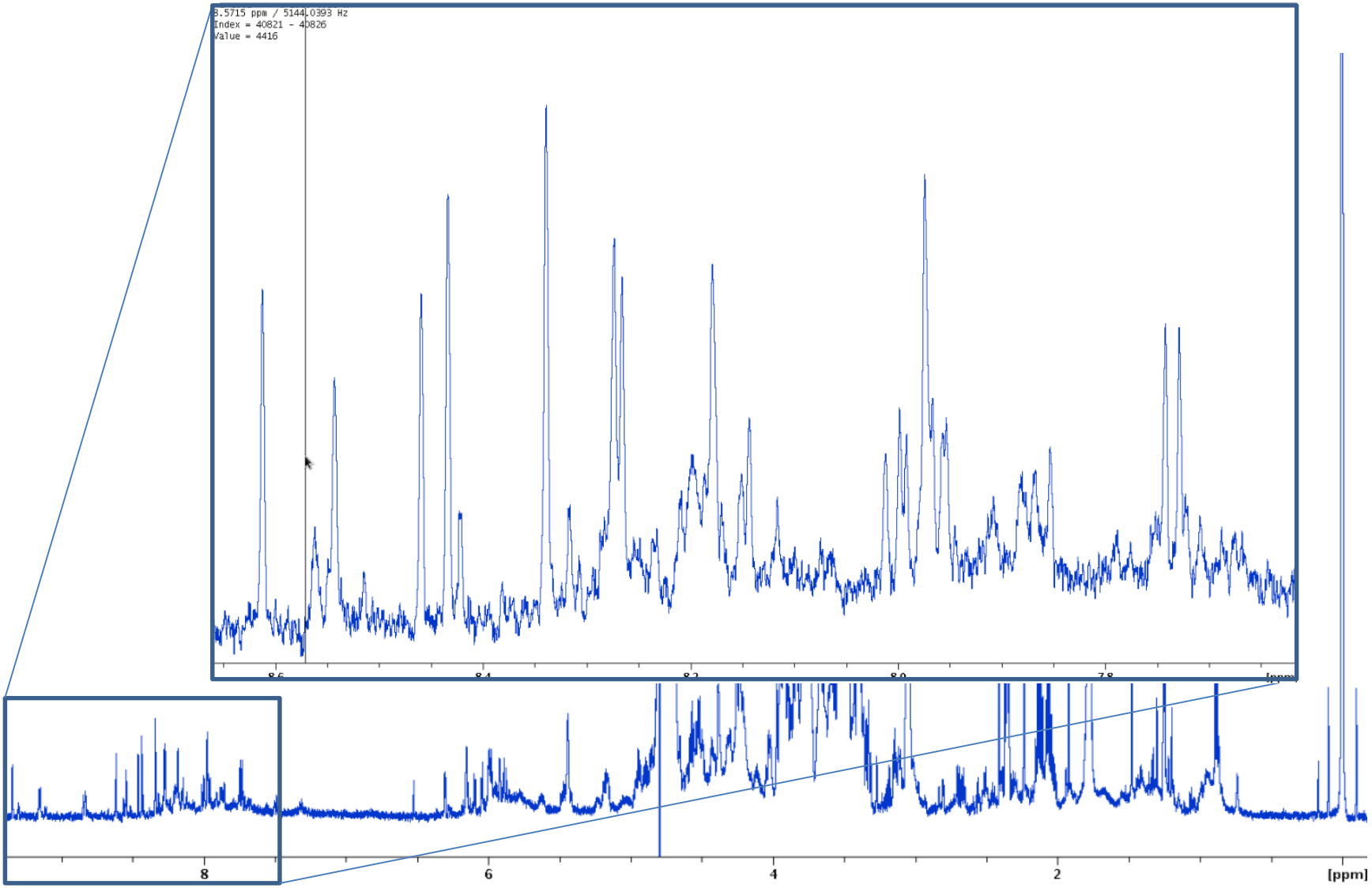
Limitations:

- Time consuming
- Fiddly (need to be precise)
- Need to assign metabolite to peaks to get the maximum benefit.

Preparing a Pattern File

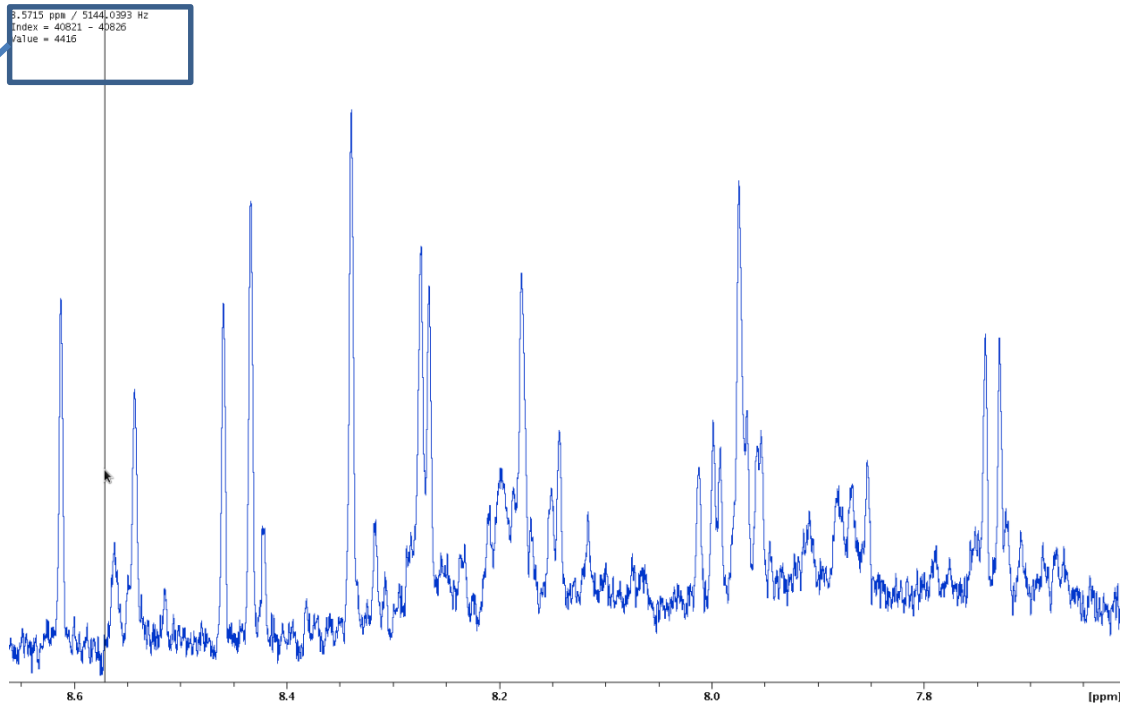


Preparing a Pattern File



Preparing a Pattern File

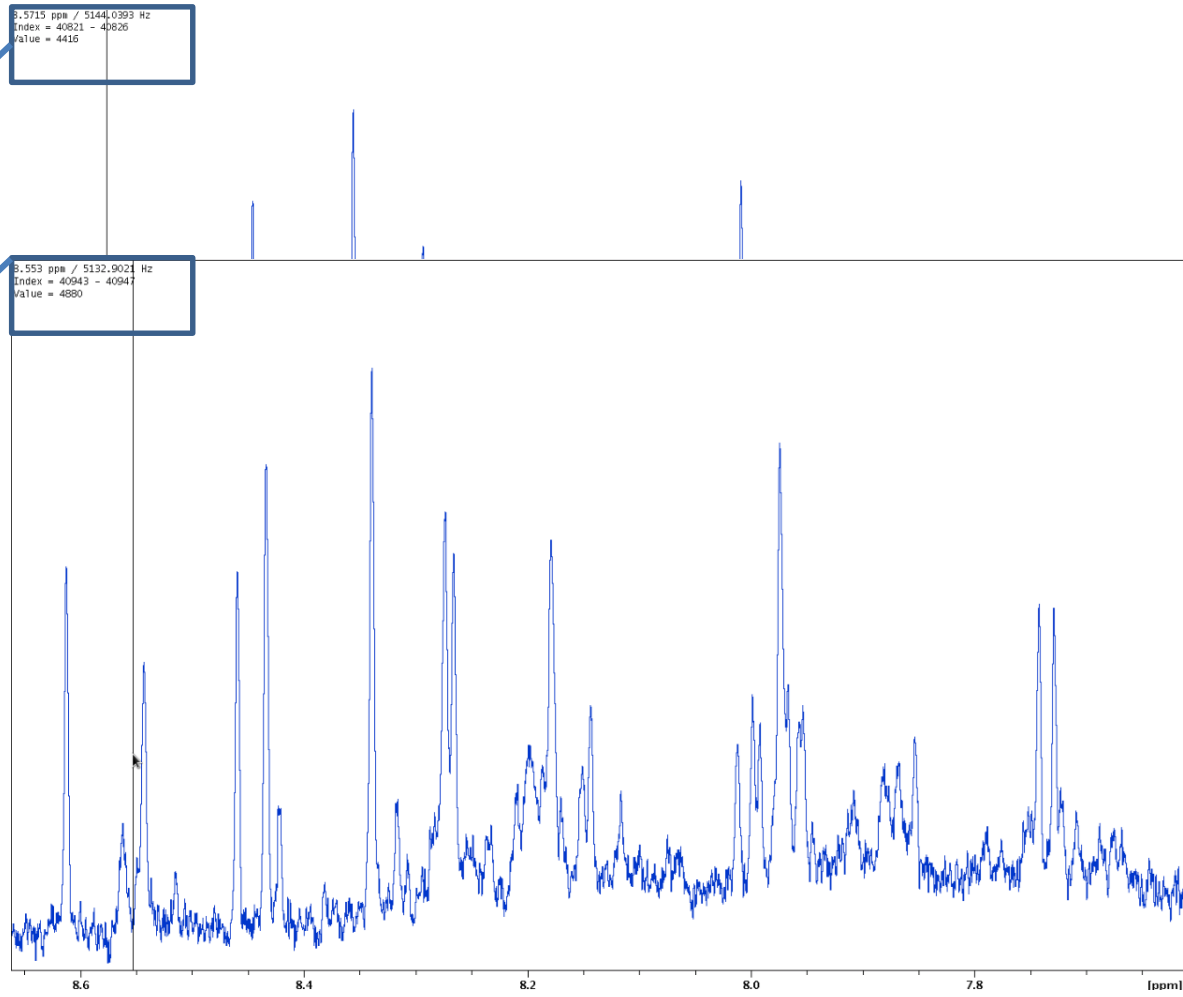
8.5715 ppm / 5144.0393 Hz
Index = 40821 - 40826
Value = 4416



Preparing a Pattern File

8.5715 ppm / 5144.0393 Hz
Index = 40821 - 40826
Value = 4416

8.553 ppm / 5132.9021 Hz
Index = 40943 - 40947
Value = 4880

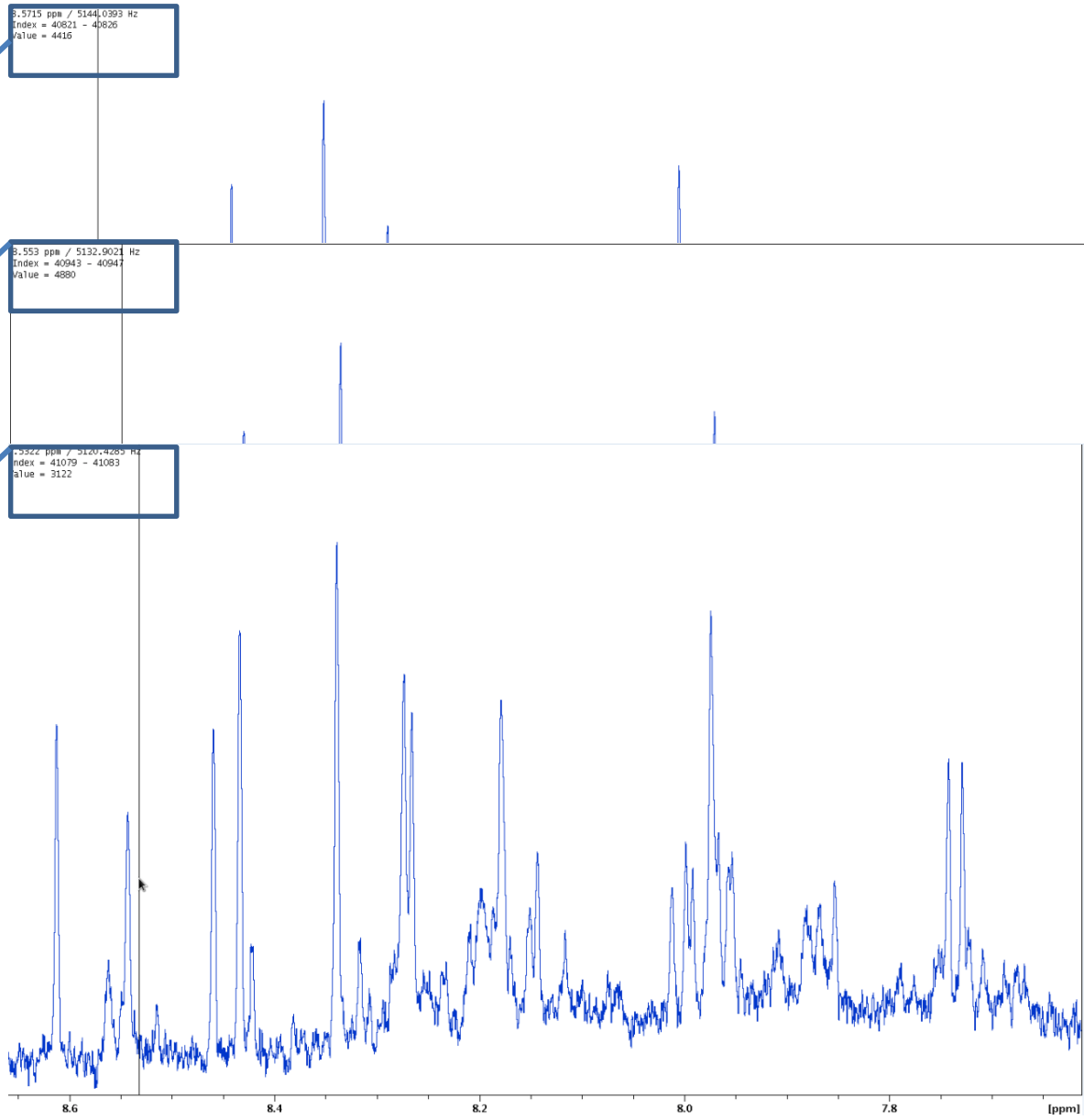


Preparing a Pattern File

8.5715 ppm / 5144.0393 Hz
Index = 40821 - 40826
Value = 4416

8.553 ppm / 5132.9021 Hz
Index = 40943 - 40947
Value = 4880

8.5322 ppm / 5120.4285 Hz
Index = 41079 - 41083
Value = 3122



Preparing a Pattern File

8.5715 ppm / 5144.0393 Hz
Index = 40821 - 40826
Value = 4416

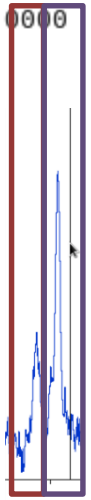
8.553 ppm / 5132.9021 Hz
Index = 40943 - 40947
Value = 4880

8.5322 ppm / 5120.4285 Hz
Index = 41079 - 41083
Value = 3122

0.0000	0.0000	8.270	8.259	0 AMP
0.0000	0.0000	8.279	8.270	0 ADP
0.0000	0.0000	8.344	8.330	0 aro11
0.0000	0.0000	8.389	8.378	0 aro12
0.0000	0.0000	8.427	8.417	0 NADP2
0.0000	0.0000	8.445	8.433	0 NAD2
0.0000	0.0000	8.464	8.454	0 formate
0.0000	0.0000	8.505	8.496	0 aro13
0.0000	0.0000	8.522	8.505	0 ATP
0.0000	0.0000	8.553	8.532	0 ADP
0.0000	0.0000	8.572	8.553	0 aro14
0.0000	0.0000	8.617	8.596	0 AMP
0.0000	0.0000	8.858	8.822	0 NADP+NAD8
0.0000	0.0000	9.125	9.000	0 NADP3
0.0000	0.0000	9.175	9.137	0 NAD9
0.0000	0.0000	9.301	9.291	0 NADP4
0.0000	0.0000	9.355	9.335	0 NAD10

aro14

ADP



Pattern File Format

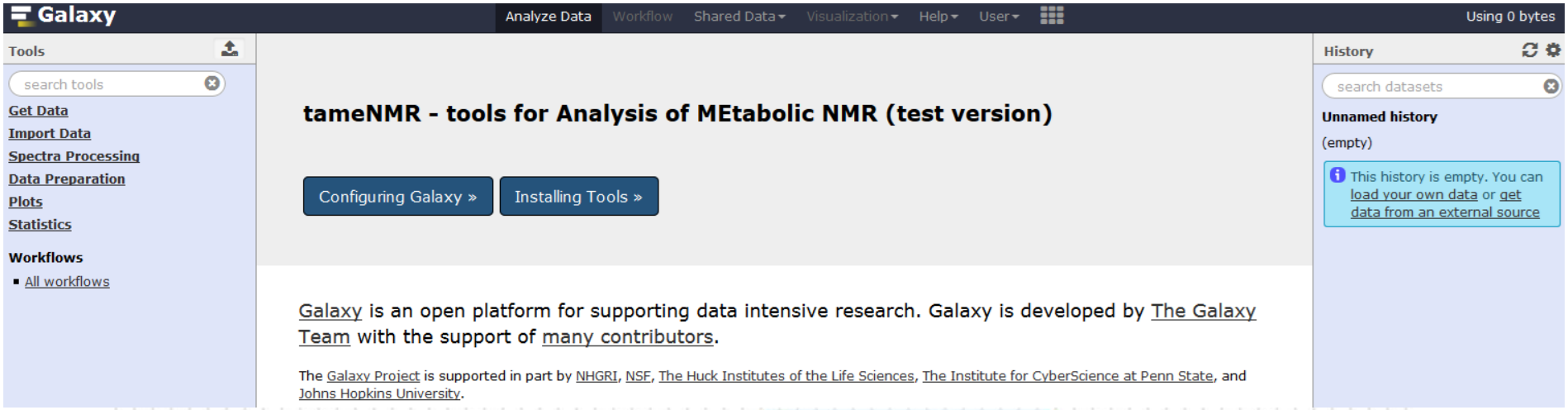
#	Data
1	$##DATE The date in ddmmyy format e.g 150915
2	
3	PATTERN = Sample name and the type of metabolites
4	GROUP = The group you are working for
5	DESCRIPTION = Any additional information
6	AUTHOR = Your name
7	DIM = 2 Leave it as 2
8	ORIGIN = 1 Leave it as 1
9	ITEMS = 11 Write the exact number of peaks you selected*
10	0.0000 0.0000 0.6834 0.6751 0 Chol-18
11	0.0000 0.0000 1.0128 1.0064 0 Chol-19
12	0.0000 0.0000 1.2698 1.2362 0 TAA ?
13	0.0000 0.0000 2.3673 2.3608 0 TAA-1
14	0.0000 0.0000 2.4839 2.4581 0 =CH-CH2-CH=?
15	0.0000 0.0000 2.8184 2.7795 0 TAA-2
16	0.0000 0.0000 3.4995 3.4868 0 -N+(CH3)3
17	0.0000 0.0000 4.0191 4.0084 0 glyceryl-C3H (PC/PE)
18	0.0000 0.0000 4.1217 4.1115 0 LDL
19	0.0000 0.0000 4.1319 4.1217 0 glyceryl backbone CH2-1
20	0.0000 0.0000 4.2554 4.2461 0 glyceryl backbone CH2-2

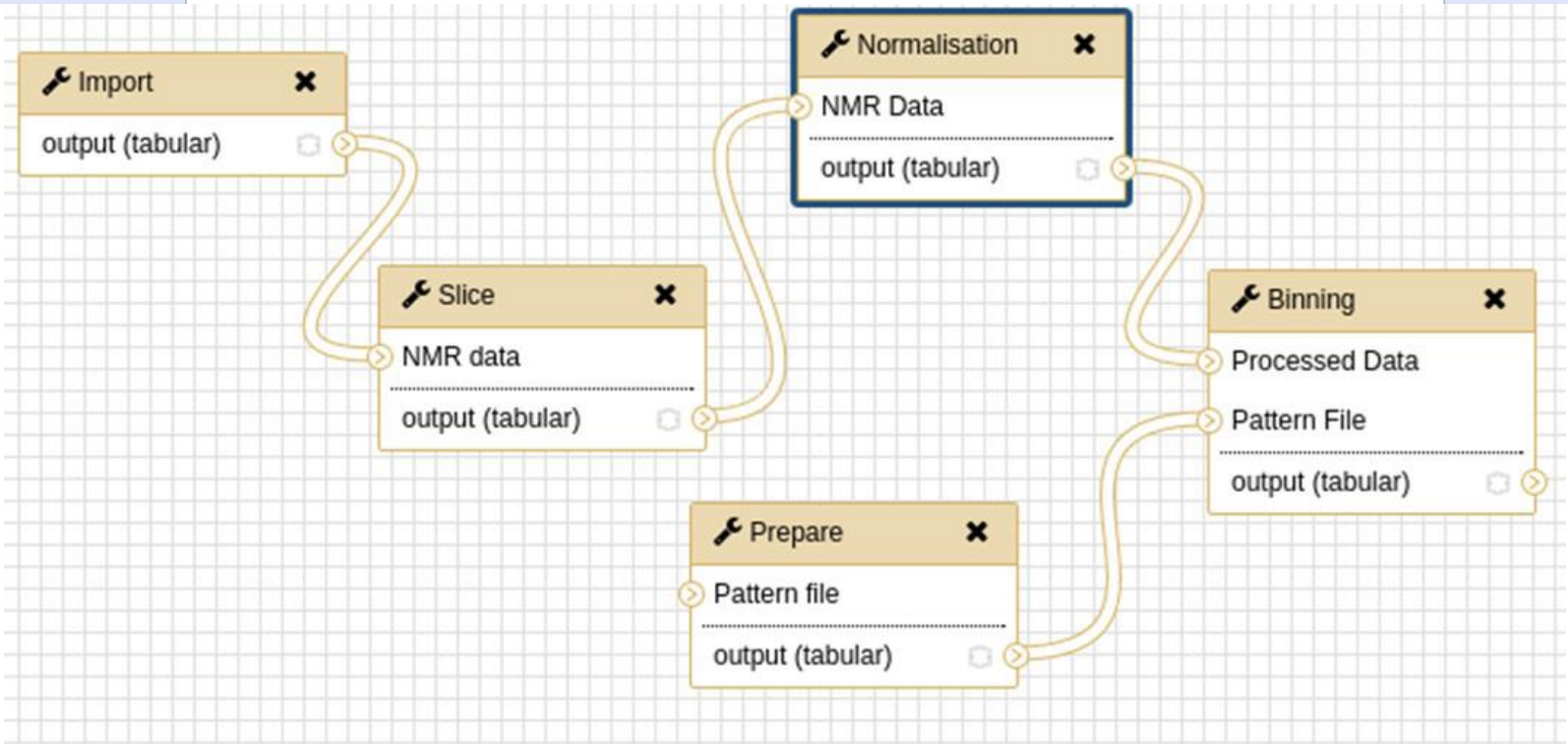
Pattern File Format

#	Data	
1	$##DATE 150915	*The date in ddmmyy format*
2		
3	PATTERN = Plasma Antibiotics	*Sample name and the type of metabolites*
4	GROUP = IGH	*The group you are working for*
5	DESCRIPTION = Control	*Any additional information*
6	AUTHOR = A Researcher	*Your name*
7	DIM = 2	*Leave it as 2*
8	ORIGIN = 1	*Leave it as 1*
9	ITEMS = 11	*Write the exact number of peaks you selected**
10	0.0000 0.0000 0.6834 0.6751 0 Chol-18	
11	0.0000 0.0000 1.0128 1.0064 0 Chol-19	
12	0.0000 0.0000 1.2698 1.2362 0 TAA ?	
13	0.0000 0.0000 2.3673 2.3608 0 TAA-1	
14	0.0000 0.0000 2.4839 2.4581 0 =CH-CH2-CH=?	
15	0.0000 0.0000 2.8184 2.7795 0 TAA-2	
16	0.0000 0.0000 3.4995 3.4868 0 -N+(CH3)3	
17	0.0000 0.0000 4.0191 4.0084 0 glyceryl-C3H (PC/PE)	
18	0.0000 0.0000 4.1217 4.1115 0 LDL	
19	0.0000 0.0000 4.1319 4.1217 0 glyceryl backbone CH2-1	
20	0.0000 0.0000 4.2554 4.2461 0 glyceryl backbone CH2-2	

- Always use plain text.
- Never use a word processor (e.g MS Word, OpenOffice, etc).
- If prepared on Windows always check you file on linux text editor (e.g nedit). Files prepared on Windows may introduce “<cr>” in your line. If this happens simply delete every <cr> and save the file before you proceed.

Galaxy - TameNMR





Galaxy - TameNMR

1. Import - Import of NMR data (from Bruker raw files)
- Bruker to csv

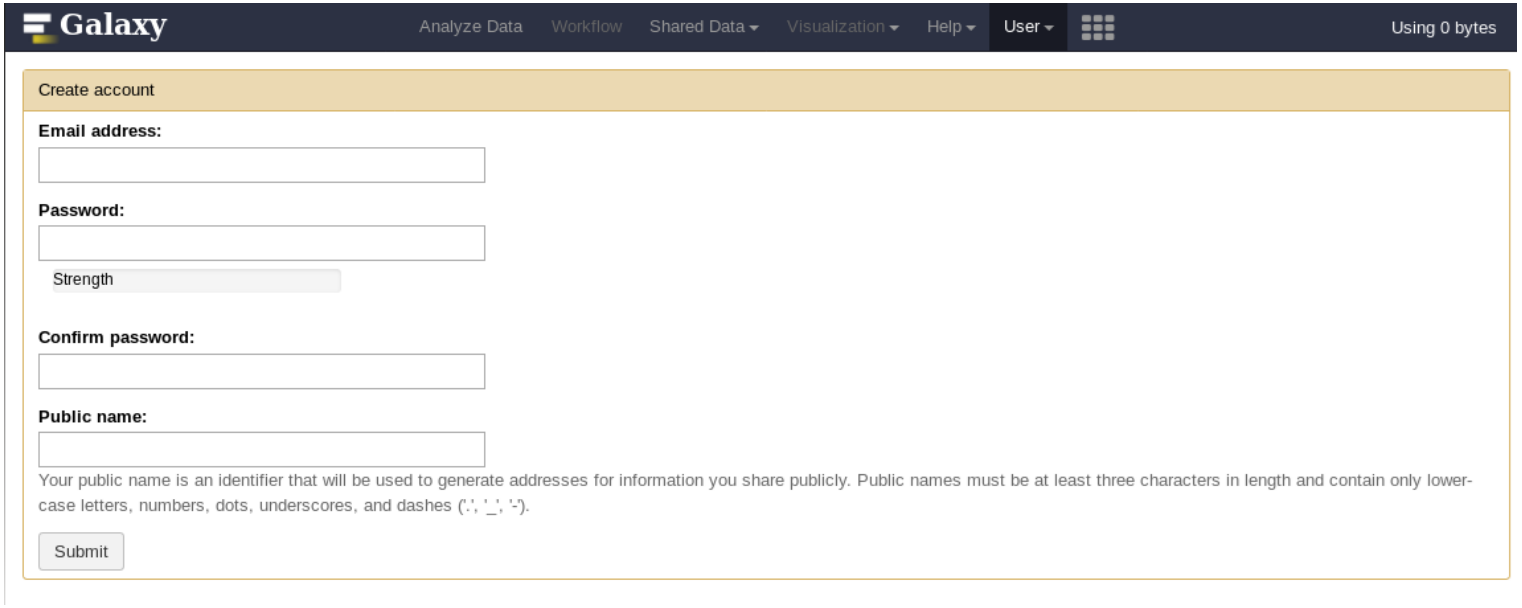
2. Process Spectra
- Normalisation
- Peak Picking
- Spectra alignment
- Binning

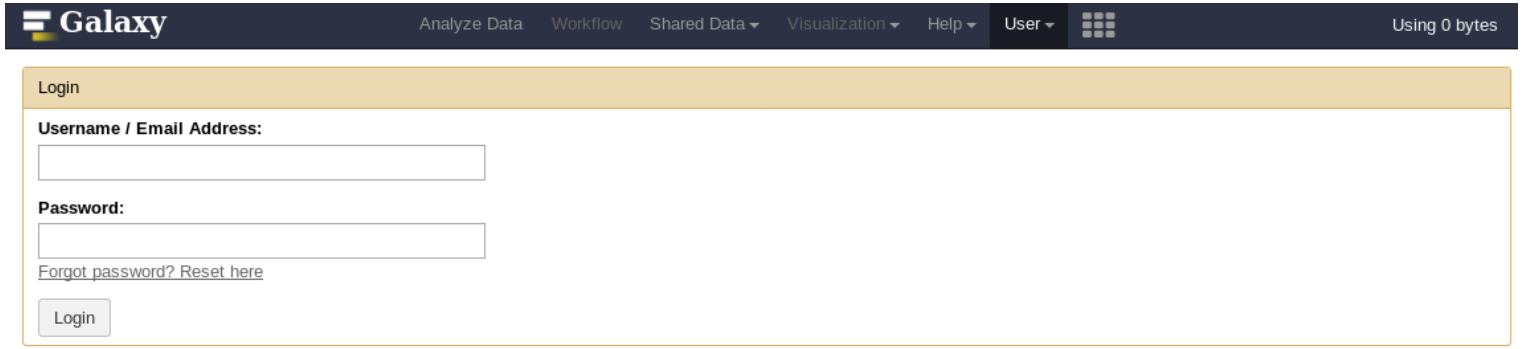
3. Process Data
- Scaling
- Make factor template (for grouping observations)

4. Statistics - univariate and multivariate statistics
- t-tests
- one-way ANOVA
- Principal component analysis (PCA)
- Partial least squares discriminant analysis (PLS-DA)

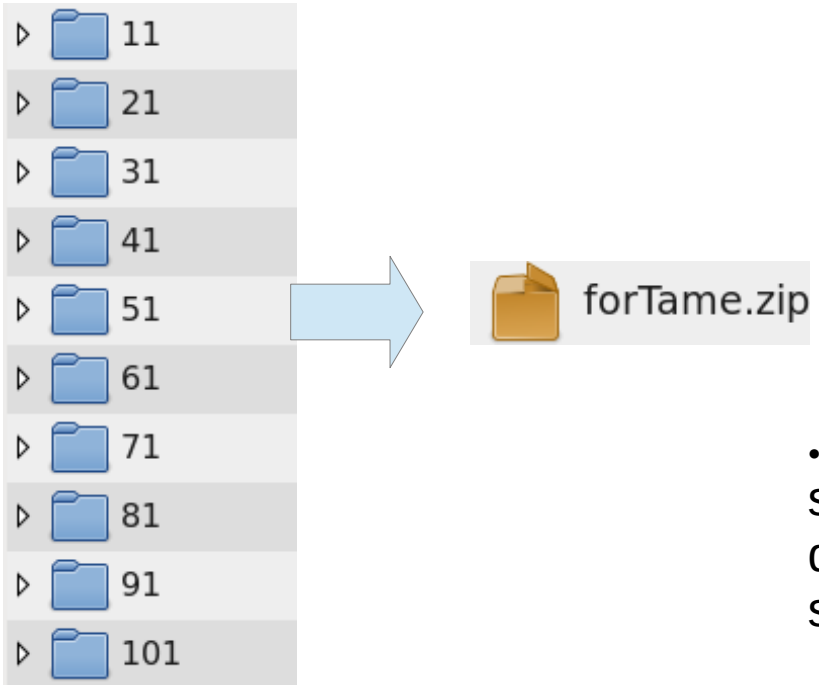
5. Plots - various plotting tools for:
- Raw NMR spectra
- Quantiles of spectra
- Significant bins (*p*-values from t-tests of ANOVA)
- Significant bins (mean value comparison)

TameNMR – Register & Login





TameNMR – Data Formats

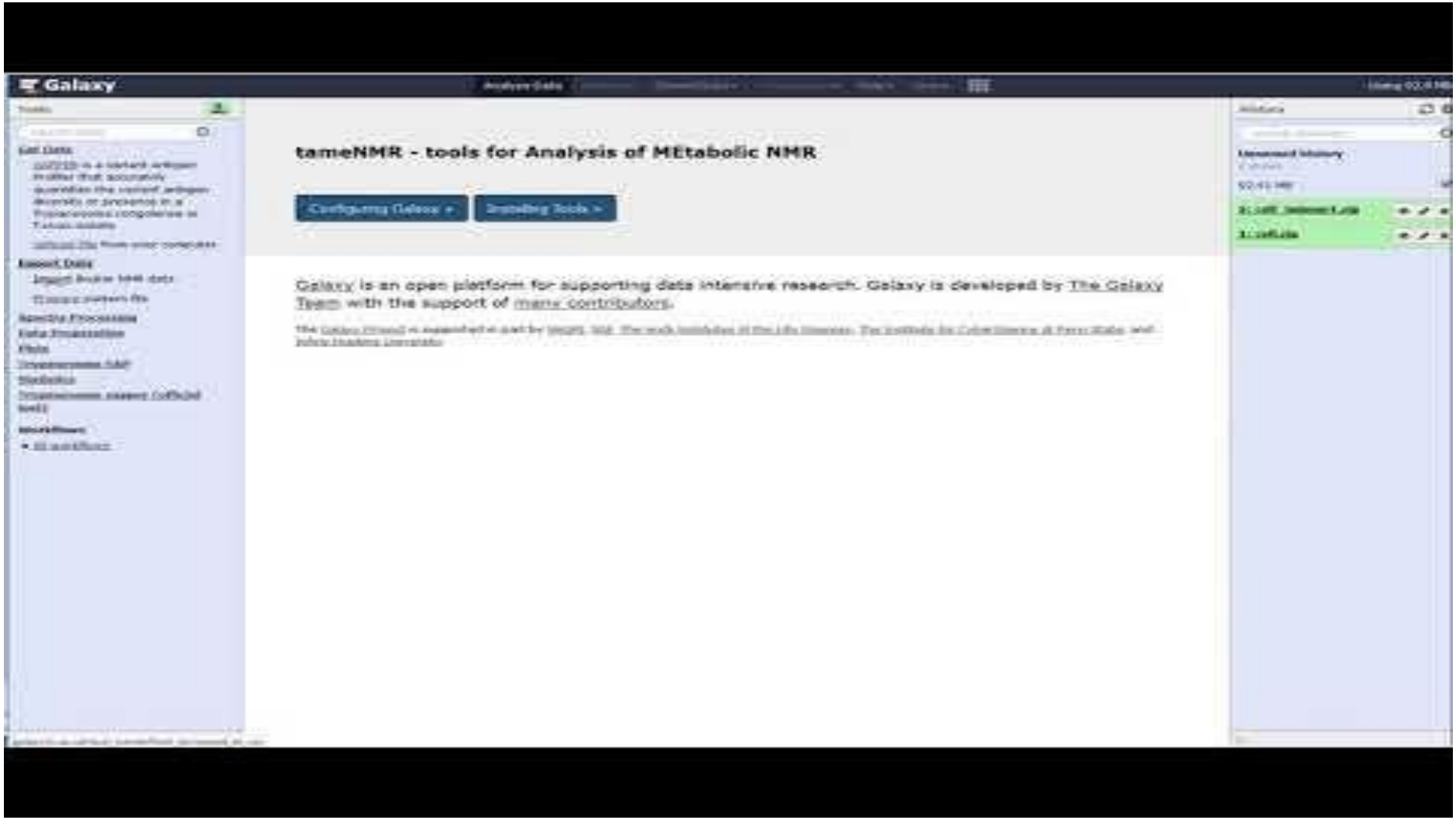


- A set of Bruker format Spectra can be zipped together for upload (do not zip dataset parent folder but individual experiment numbers)

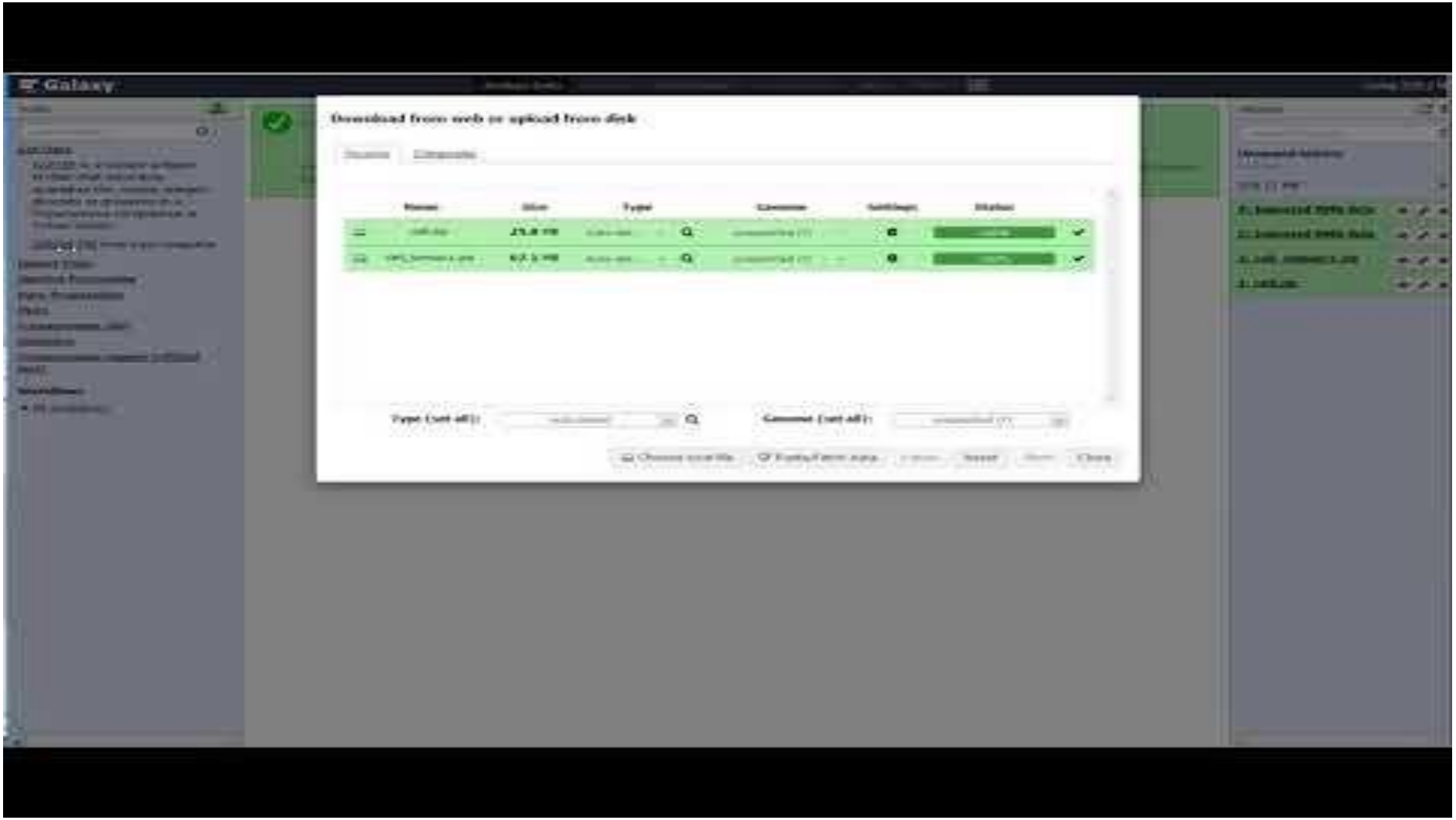
- Alternatively prepare a table with tab separated columns (every spectrum in a different column, header containing spectra ID)

- Tab separated column formats must also be prepared for bin boundary **pattern** files (peak annotation) and sample groupings (for statistical analysis).

1	2	3	4
"ppm"	"A194"	"B195"	"C196"
9.99990639748	583.224970898215	49.6107431222747	-219.231228317496
9.99977543254	419.501182307945	-223.745293769119	-79.6887423671469
9.9996444676	357.159005369531	-370.219117690402	94.052768959037
9.99951350265	521.642853750682	-461.051418850098	502.15241976322
9.99938253771	605.491267953802	-396.936482237745	884.510165552195
9.99925157277	490.273113290677	-151.359092344189	928.059976069032
9.99912060783	403.868134336411	32.781835693021	768.74788144062
9.99898964288	383.933838911936	-27.7112639851076	710.943027841694
9.99885867794	253.859970154316	-194.36630021895	755.648063562432

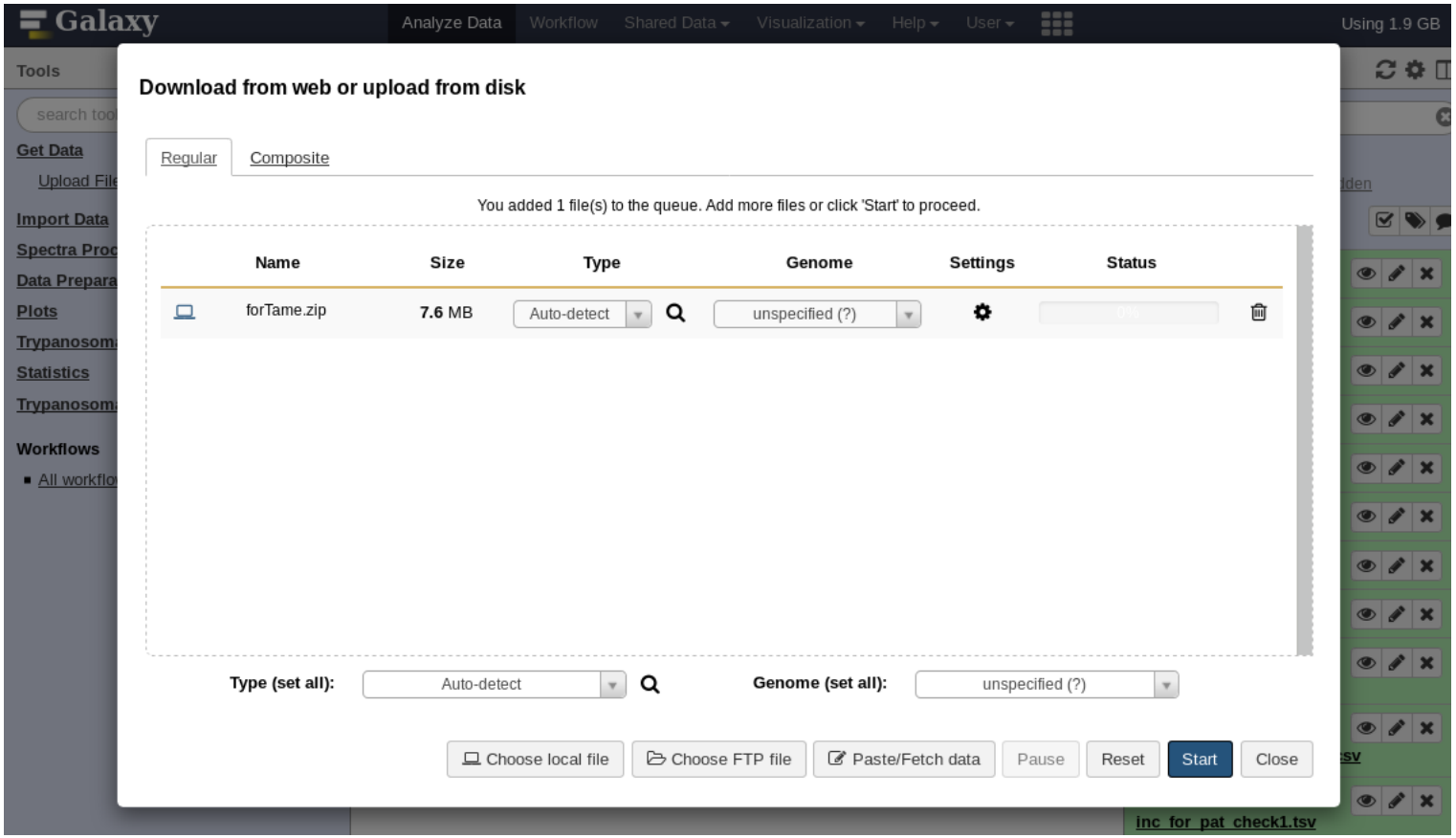


Import Spectra – Bruker data:
https://youtu.be/Fc2g5oIF6Jw



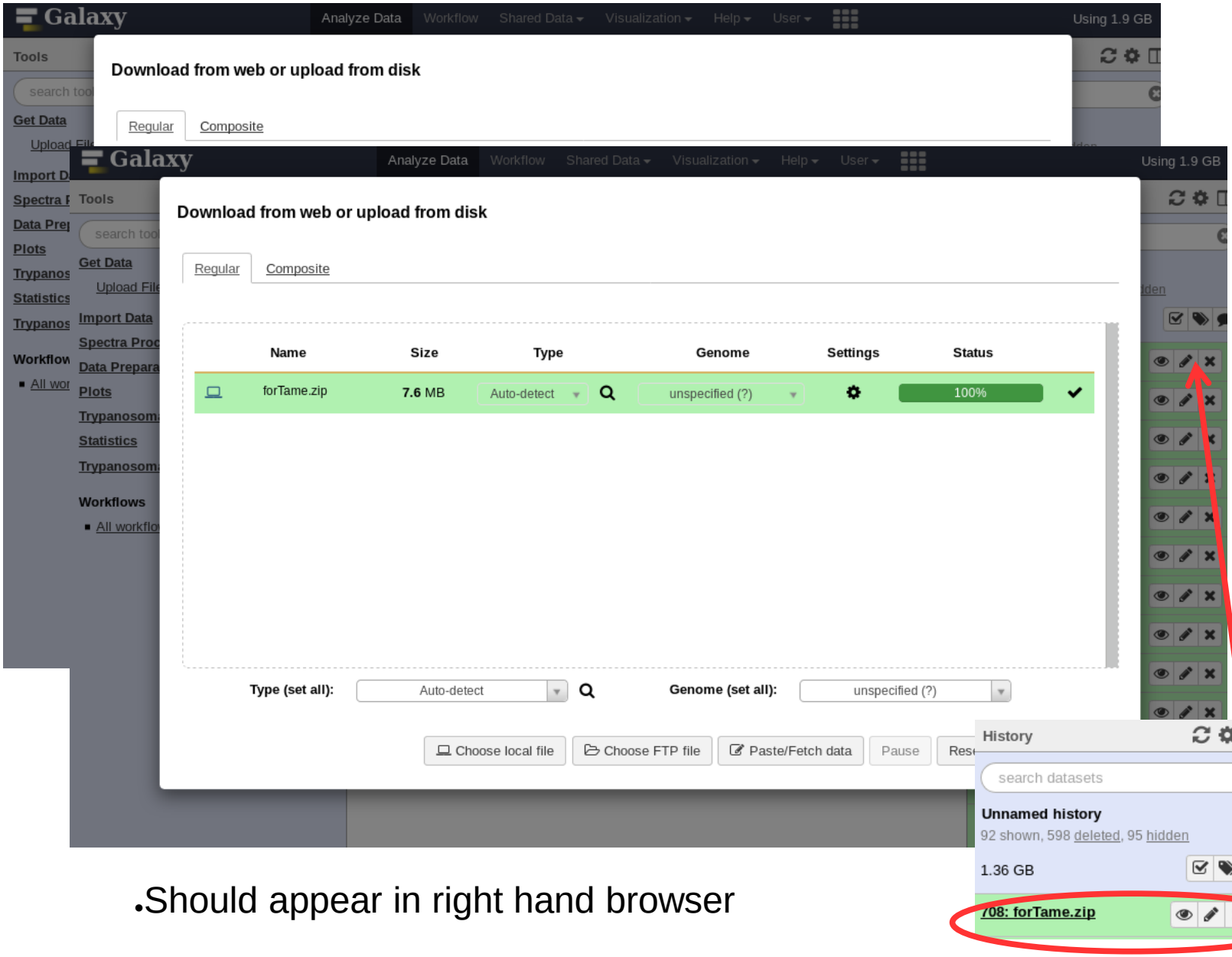
Import peak boundaries – pattern file:
https://youtu.be/wH2g4CDa2Dk

TameNMR – Data import



- Upload spectra and annotation *via* drag and drop

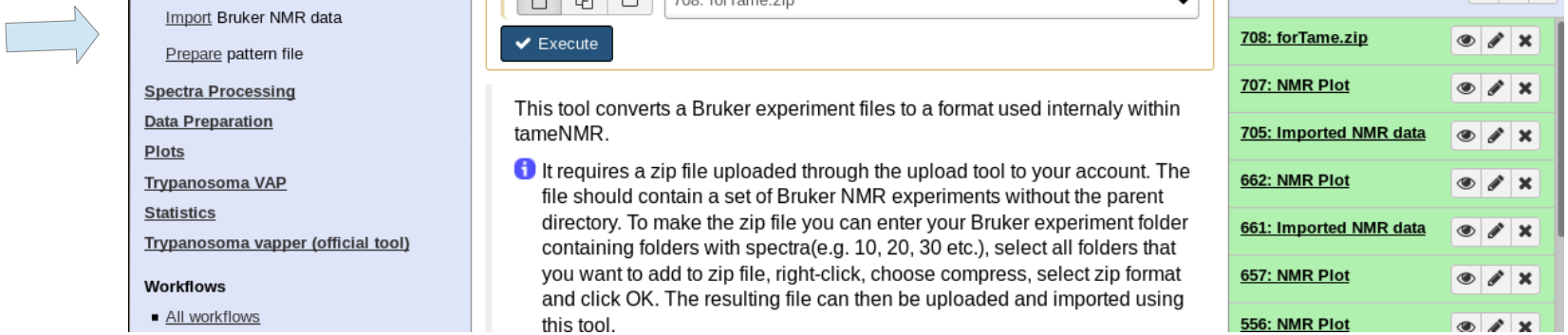
TameNMR – Data import



- Should appear in right hand browser

TameNMR – Data import

```
0.0000   0.0000  8.270     8.259        0 AMP
0.0000   0.0000  8.279     8.270        0 ADP
0.0000   0.0000  8.344     8.330        0 aro11
0.0000   0.0000  8.389     8.378        0 aro12
0.0000   0.0000  8.427     8.417        0 NADP2
0.0000   0.0000  8.445     8.433        0 NAD2
0.0000   0.0000  8.464     8.454        0 formate
0.0000   0.0000  8.505     8.496        0 aro13
0.0000   0.0000  8.522     8.505        0 ATP
0.0000   0.0000  8.549     8.537        0 ADP
0.0000   0.0000  8.570     8.560        0 aro14
0.0000   0.0000  8.617     8.596        0 AMP
0.0000   0.0000  8.858     8.822        0 NADP+NAD8
0.0000   0.
0.0000   0.
0.0000   0.
0.0000   0.
```

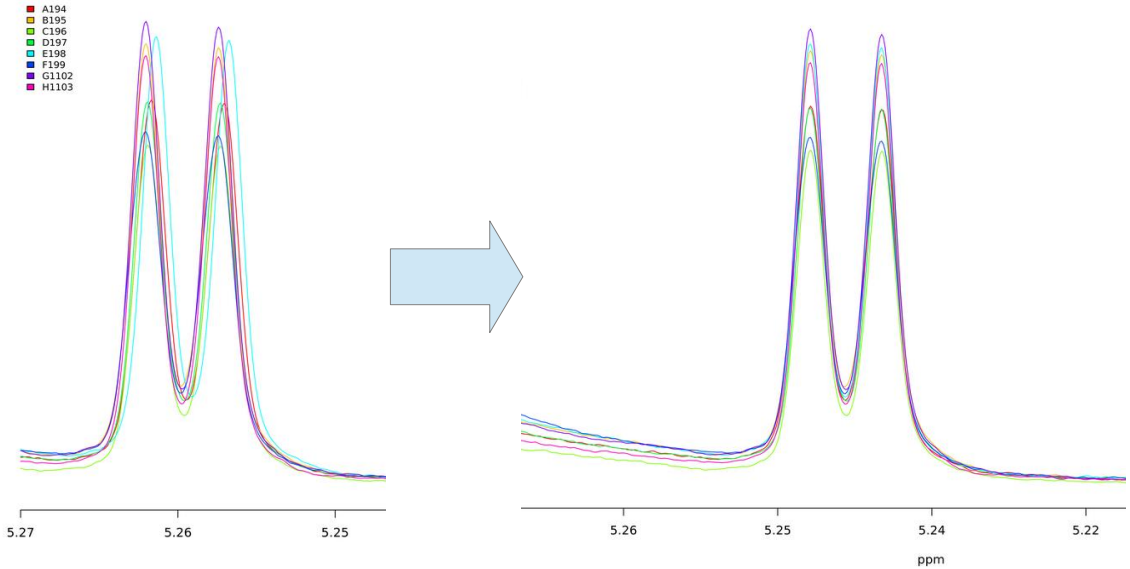


oaded

- Import the bruker data to unzip and tabulate (column per spectrum).
- Import from spreadsheet/text file the **Pattern** or bin boundaries

TameNMR – Process Spectra

Alignment

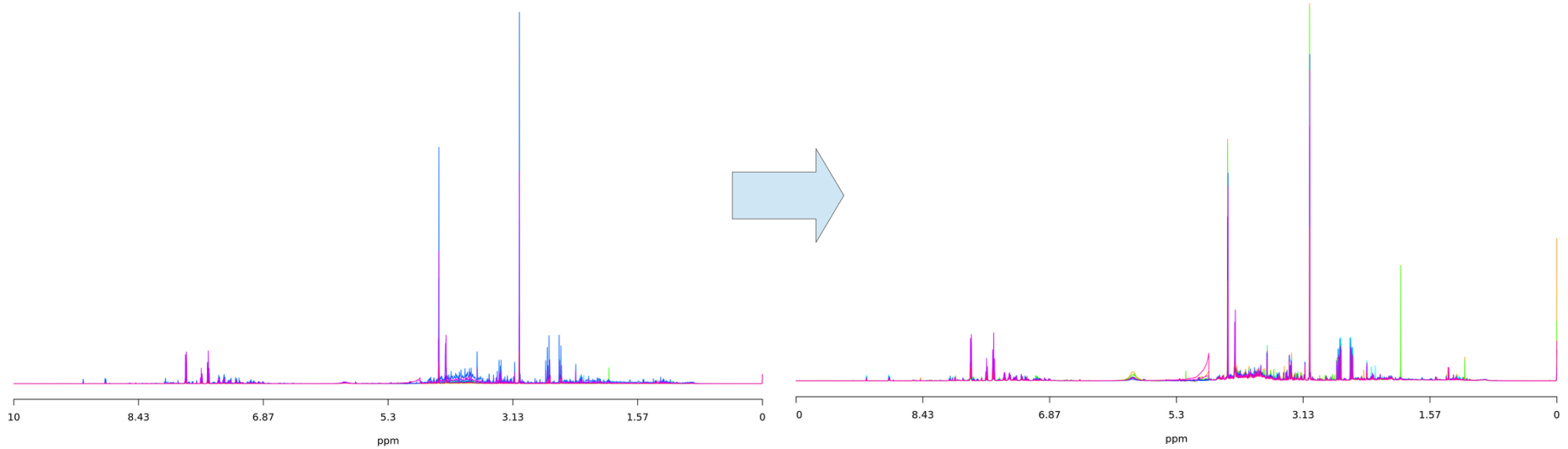


Align to different reference pea

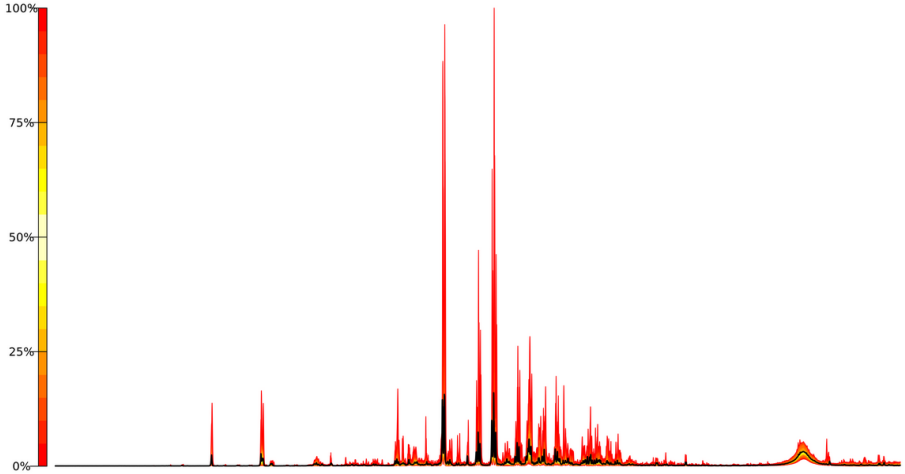
Data normalisation may be app

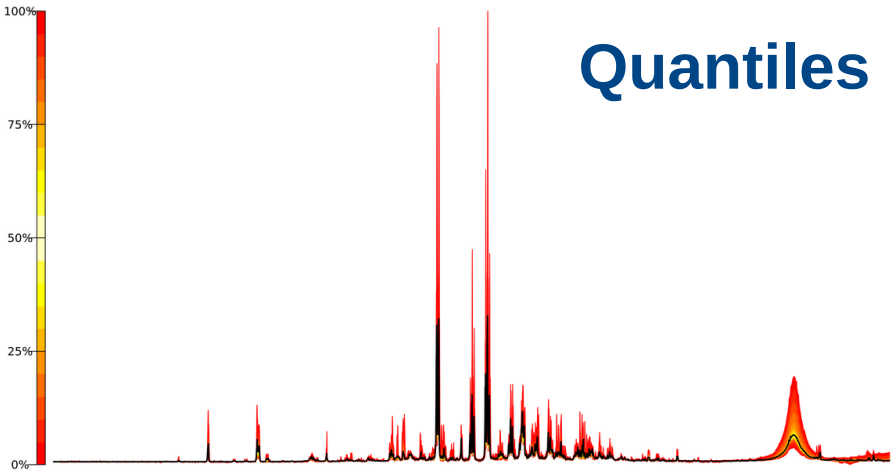
- Peak picking function enables
- Binning spectra using predefi

Normalisation



TameNMR – Process Spectra

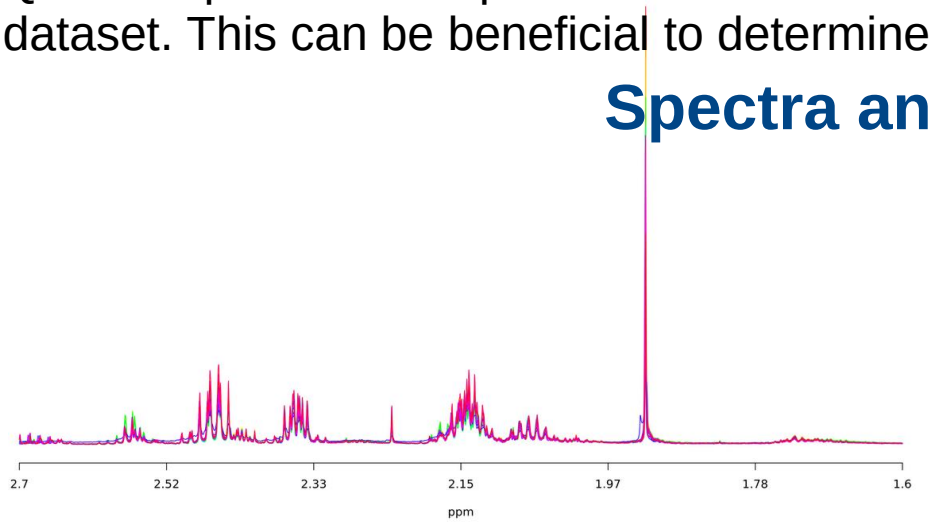


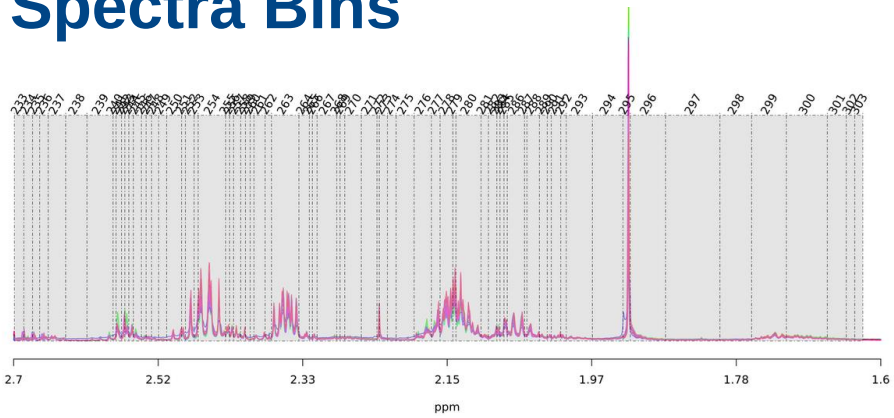


Quantiles

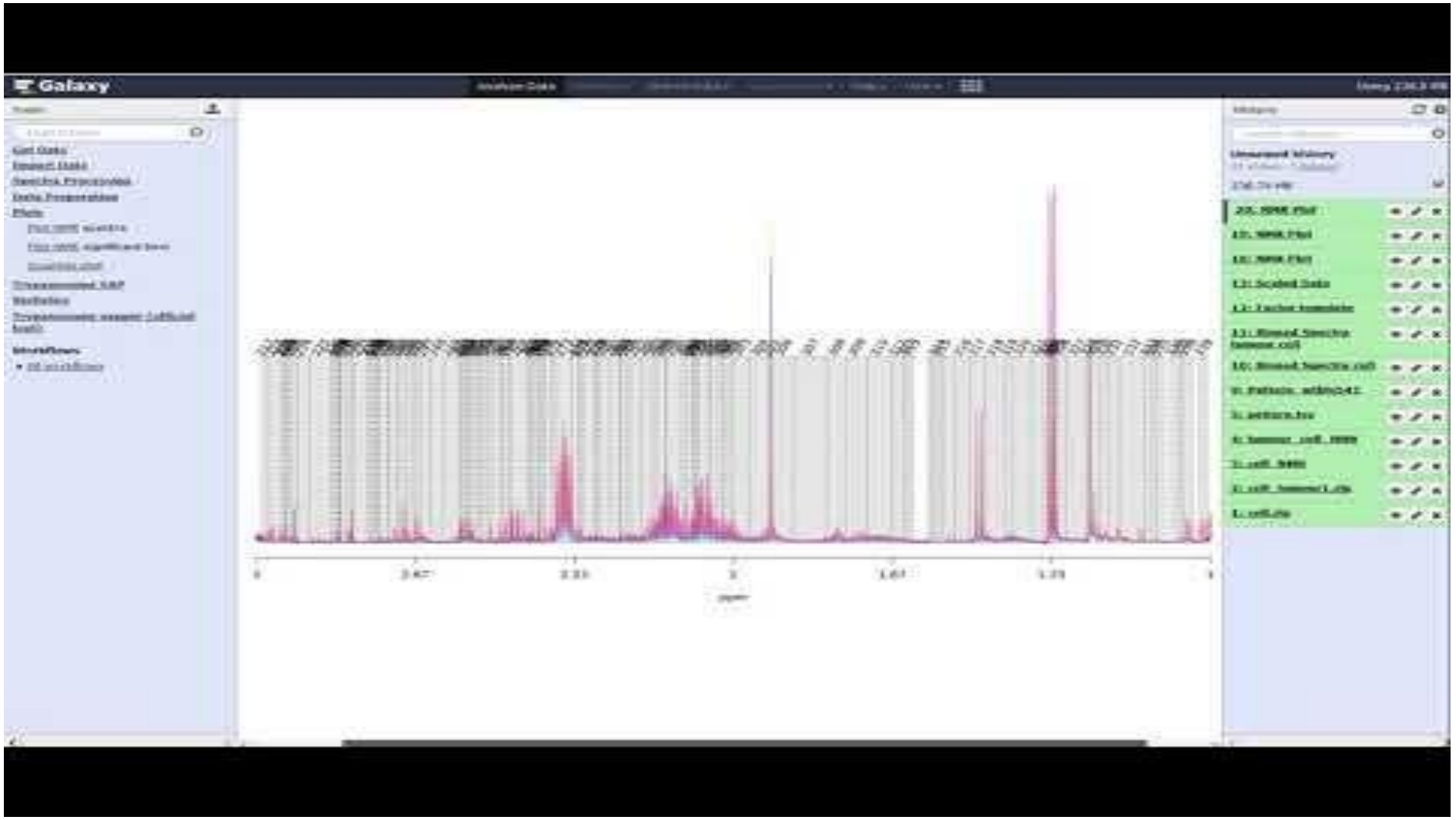
Quantile plot of the spectra can show median (black) and variance for a particular dataset. This can be beneficial to determine normalisation efficacy

Spectra and Spectra Bins



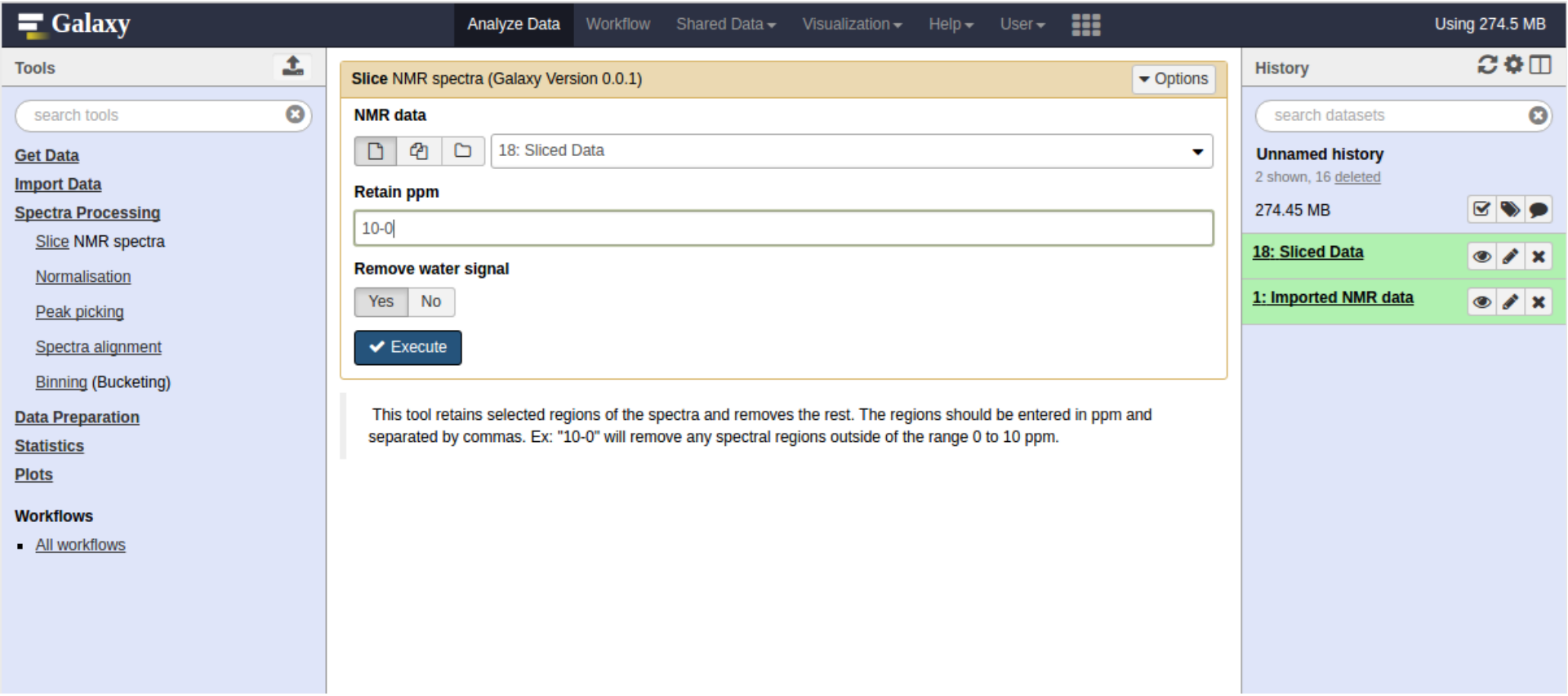


Quality of overlay of bins (peak boundaries) with spectra is key to quality of data analysis. This function was a driving factor for generating this toolkit.

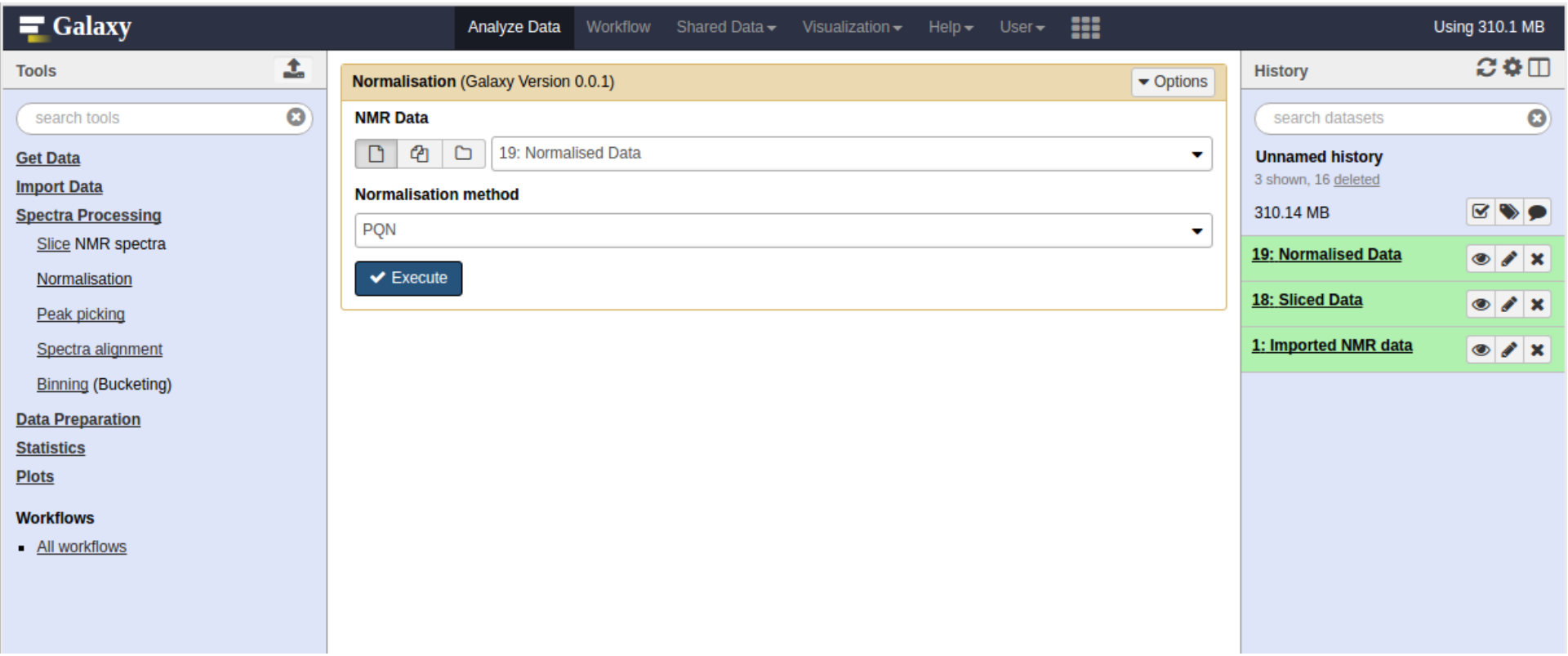


Visualise data – Spectra and Annotation:
https://youtu.be/2nO1AJvx-aY

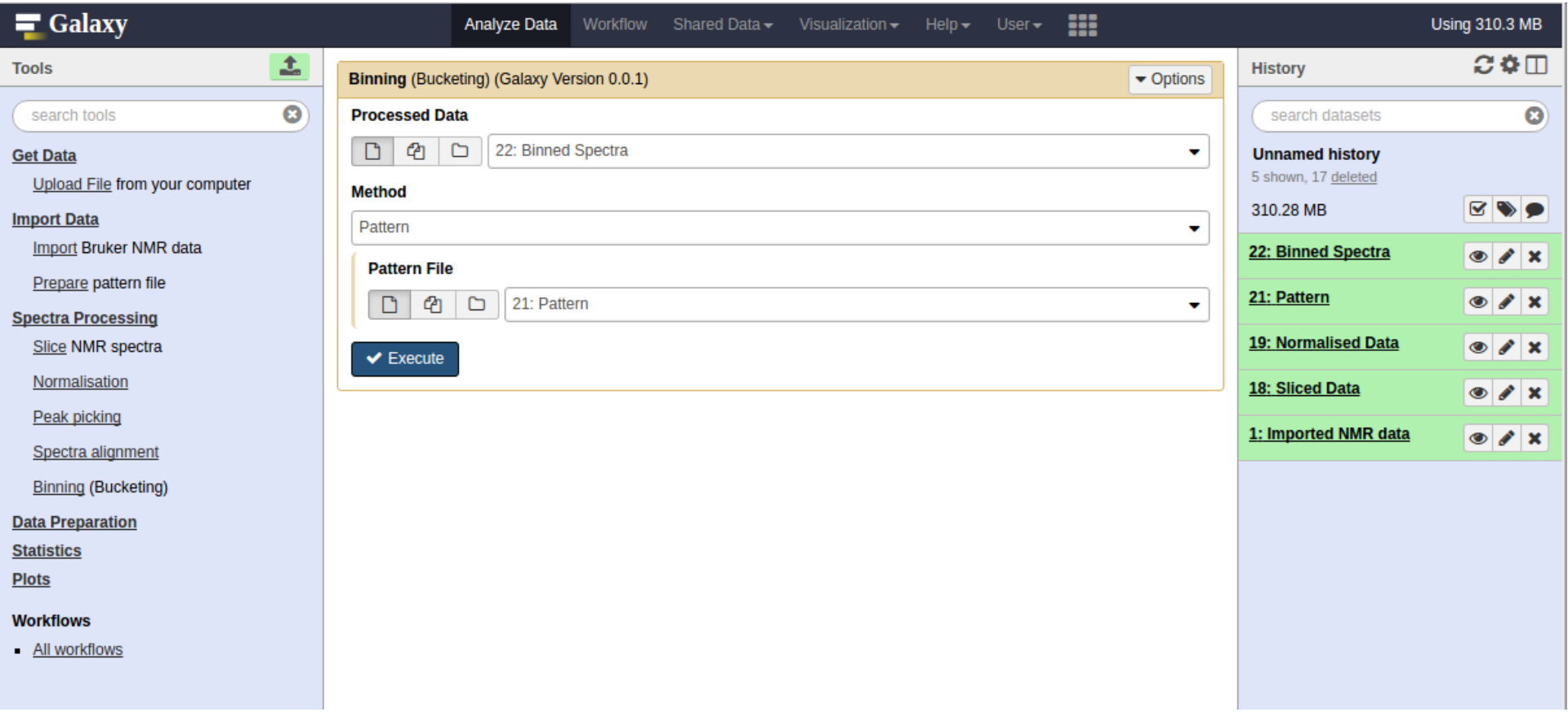
TameNMR – spectra slicing



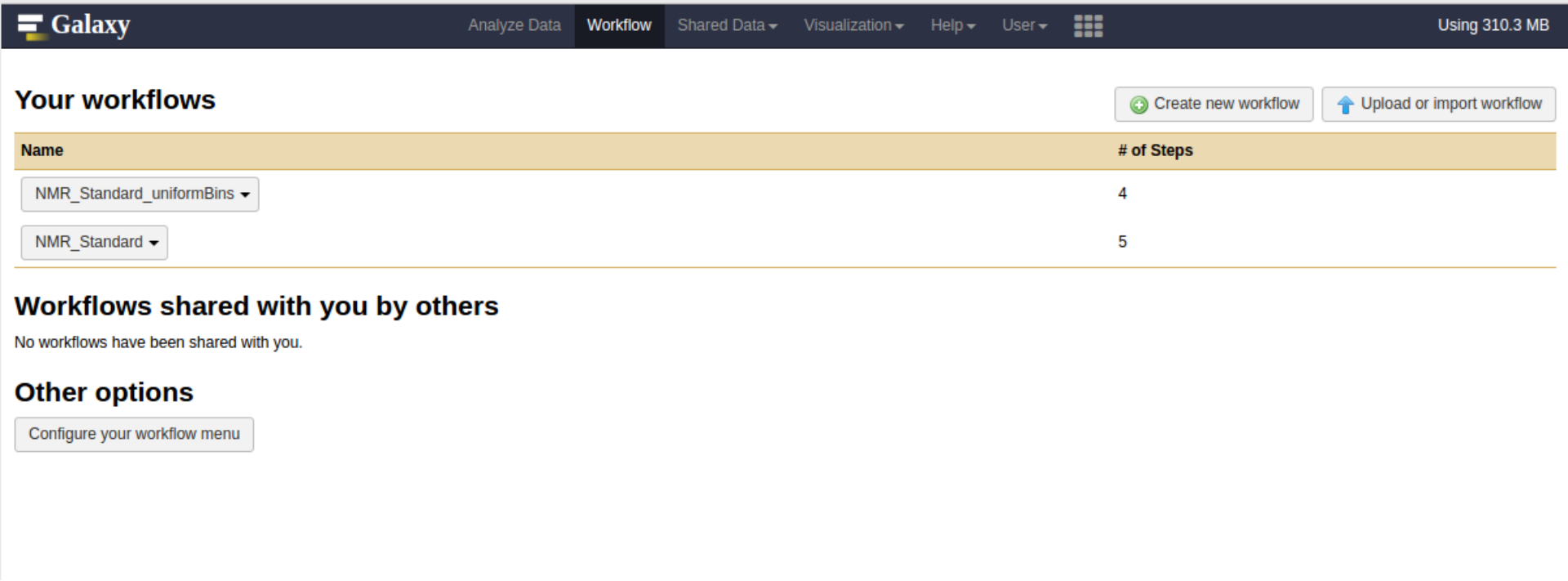
TameNMR - Normalisation



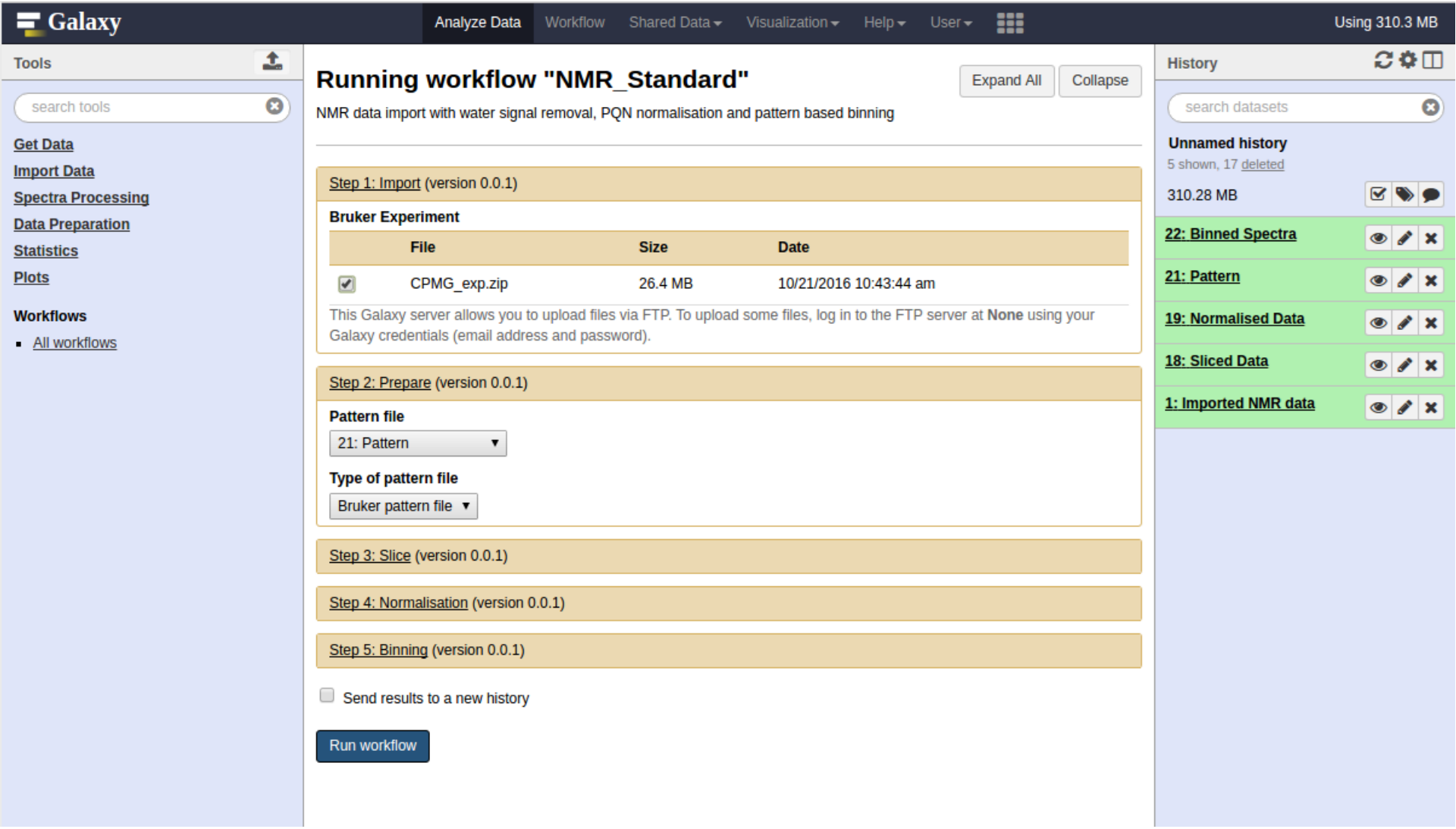
TameNMR – Binning



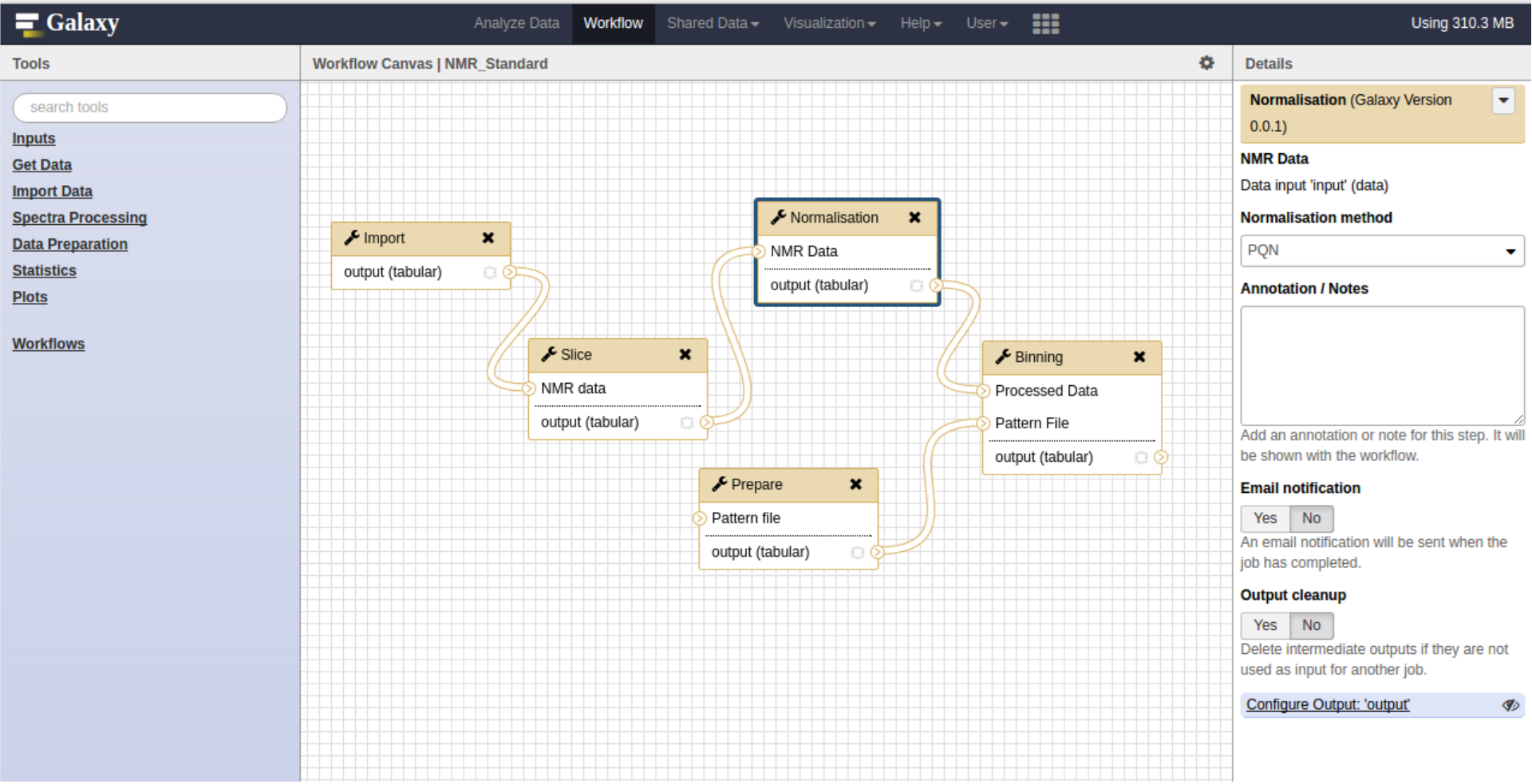
TameNMR - Workflows



TameNMR - Workflows



TameNMR - workflows



Acknowledgements

Liverpool Metabolomics beta testers:

Rudi Grosman
Dr Eva Caamaño-Gutiérrez (CBF)
Adika Sen
Michelle Tan
Dan Sadler
Nicola Beesley
Alan Reynolds
Yetunde Adegbite
Thomas Leather
Zain Ghanameh
Mattia Scalibrin
Brendan Norman
Andrew Davison

Technology Directorate:

Dr Duncan Robertson
Julie Boileau

Liverpool Computational Biology (CBF):

Prof Andy Jones
Dr Arturas Grauslys

Funding:

BBSRC TDRF2 call (BB/M020282/1)

@LivUniNMR

@LivUniCBF

E: mphelan@liverpool.ac.uk

Download from git-hub:
https://github.com/PGB-LIV/tameNMR

tameNMR Protocols/guide:
https://sites.google.com/view/tamenmr

tameNMR mailing list:
https://forms.gle/q8VS2rHT479dNow27

UNIVERSITY OF LIVERPOOL

NMR Metabolomics

CBF
COMPUTATIONAL BIOLOGY FACILITY

@LivUniNMR

@LivUniCBF

BBSRC
bioscience for the future

BB/M020282/1

tameNMR tutorials:
https://youtube.com/channel/UCw92VXikB44MVklWW24p2fA

CCPN tutorials:
https://youtube.com/channel/UCKyPsbC0mgaXAn03Owah50w

tameNMR mailing list:
https://forms.gle/XKfiVPuT1RiURQ7a7