EI-MAVEN Documentation

Release v0.4.1

Elucidata Corporation

Oct 18, 2018

Contents

Cont	
	El-MAVEN Features
1.2	Download
	Build
1.4	User Documentation
1.5	Contributors
1.6	Contributing
1.7	References
1.8	Acknowledgement
1.9	Release History
1.10	Bugs and Feature Requests
1.11	Copyright and License

An intuitive, opensource LC-MS data processing engine from

Travis Build

Digital Object Identifier

CHAPTER 1

Contents

1.1 EI-MAVEN Features

MAVEN and El-MAVEN share the following features:

- Multi-file chromatographic aligner
- Peak-feature detector
- Isotope and adduct calculator
- · Formula predictor
- · Pathway visualizer
- Isotopic flux animator

El-MAVEN is faster, more robust and has more user friendly features as compared to MAVEN.

1.2 Download

El-MAVEN installers are available for Windows (7, 8, 10) and MacOS (1.5 and above). Download El-MAVEN's latest version or daily build for your preferred environment.

1.3 Build

Contributors can build El-MAVEN on Windows, Ubuntu or Mac systems by following these instructions. Users are recommended to download the installers provided here.

1.3.1 Windows

- Download MSYS2 installer and follow the installation instructions provided on their website.
- Open MSYS2 and give the following commands to set up libraries and tool chains for El-MAVEN. Reopen MSYS2 when required:

- For 64 bit

- For 32 bit

```
$ pacman --force -Sy
$ pacman --force -Syu
$ pacman --force -Su
$ pacman --force -Sy base-devel msys2-devel mingw-i686-toolchain mingw-
→i686-qt5 mingw32/mingw-i686-hdf5 mingw32/mingw-i686-netcdf mingw32/
→mingw-i686-boost msys/git
```

• Open mingw64.exe from the MSYS2 folder and give the following commands:

```
$ cd <PathToInstallationFolder> #for example: cd /c/User/Admin/Desktop
$ git clone https://github.com/ElucidataInc/ElMaven.git
$ ./run.sh
$ ./bin/El_Maven_0.x #for example: ./bin/El_Maven_0.2
```

El-MAVEN loads with two windows: one for logging the application status and another for data analysis.

1.3.2 Ubuntu

• Open the terminal and give the following commands to set up the libraries and tool chains for El-MAVEN:

El-MAVEN loads with two windows: one for logging the application status and another for data analysis.

1.3.3 MacOS

- Install Xcode from App store
- Download and Install Qt5.6

This will give you the Qt5.6.2 dmg file. Using the dmg file install Qt under the directory /Users/Your_User_Name/

• Using the terminal execute the following commands:

```
$ sudo xcodebuild -license accept
$ xcode-select --install
$ /usr/bin/ruby -e "$(curl -fsSL https://raw.githubusercontent.com/
→Homebrew/install/master/install)"
$ brew install boost
$ brew install llvm@3.7
$ brew install netcdf
$ cd ~
$ touch .profile
$ echo "PATH=/Users/$USER/Qt5.6.2/5.6/clang_64/bin/:$PATH" > .profile
$ source .profile
$ mkdir ~/maven_repo
$ cd ~/maven_repo
$ git clone https://github.com/ElucidataInc/ElMaven.git
$ cd ElMaven
$ source ~/.profile
$ qmake CONFIG+=debug -o Makefile build.pro
$ make -j4
```

1.3.4 Switching Versions

Users can switch between versions once they have compiled El-MAVEN successfully on their system. Follow these steps to pull a specific release:

- Choose the version you wish to install from the list of releases. (We recommend the version tagged "Current Release". Past releases are not stable and should be avoided)
- Find the version tag (v0.2.x, 0.1.x, etc) on the left of release notes.
- · Open your terminal and move to the installation folder
- Give the following commands in the terminal:

```
$ cd ElMaven
$ ./uninstall.sh
$ git checkout develop
$ git pull
$ git checkout v0.x.y (Example: v0.4.1)
```

- Build the new version using the following commands:
 - * For Windows and Ubuntu

\$./run.sh

* For MacOS

```
$ source ~/.profile
```

```
$ qmake CONFIG+=debug -o Makefile build.pro
```

```
$ make -j4
```

1.4 User Documentation

Welcome to the El-MAVEN user documentation!

El-MAVEN is an open source LC-MS data processing engine that is optimal for isotopomer labeling and untargeted metabolomic profiling experiments. Currently El-MAVEN exists as a desktop application that runs on Windows, Ubuntu and MacOS. The software can be used to view the mass spectra, align chromatograms, perform peak-feature detection and alignment for labeled and unlabeled mass spectrometry data. The aim of this software package is to reduce complexity of metabolomics analysis by using a highly intuitive interface for exploring and validating metabolomics data.

Find out more about the software through the links below.

1.4.1 Introduction

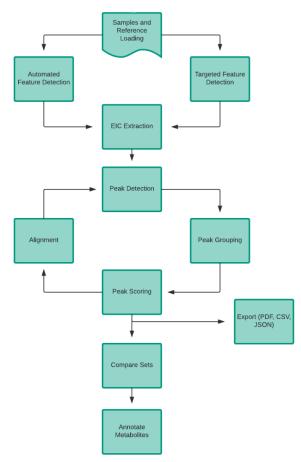
Git Repository

El-MAVEN's repository can be found on GitHub here.

1.4.2 Getting Started

EI-MAVEN User Interface

Following is the general workflow involved in El-MAVEN:



Peak detection, alignment, grouping and scoring are done multiple times for best results in the El-MAVEN workflow. Data from different cohorts can be compared using visualisation tools and easily exported to other formats.

EI-MAVEN User Interface

File Widgets Help		Auto Detect(+1) Charge 1 . AreaTop	: [m/z] MW / Compound +/- 5.000000 : ppm :
Open Align Peaks Match Options Polly	🙉 🛠 👼 🗜 🖙 👄 🐲 😰 TIC 🧮 🌉 🍫		0
Sample name filter			Widget Bar
Sample Space	EIC Window		© 80
· · · · · · · · · · · · · · · · · · ·			8

Global Settings

Global Settings can be changed from the Options dialog

Instrumentation

File Widgets Help						
Open Align Peaks Match Options Polly			Auto Detect(+1) Ch	arge 1 . AreaTop	: [m/z] MW / Compound +/	- 5.000000 ppm :
📁 🖆 🔛 🥙 🌁 🌾 😑	🙉 😭 🔂 🦨 泽 👄	🥟 🔟 TIC 📃 🎫 🍫				
Sample name filter						₩ ₩ ₩ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @
	😣 🗉 Options					Q
	Instrumentation File Import Peak	Detection Peak Filtering Isotope D	etection EIC (XIC) [BETA] Peak Grou	iping Group Rank Libraries Ex	er 🕐	
	Instrumentation Type					69
	Polarity / Ionization Mode:	Auto Detect	Q1 Accuracy	0.50 amu	1	10
	Ionization Type	ESI	Q3 Accuracy	0.50 amu		
	Instrument Type:	Full Spectrum	Filterline	All	4	
		Use Defau	lt Parameters			
					-	
······································						

- *Polarity/Ionization mode*: Polarity information is required for m/z calculation. Users can set the polarity of the metabolites in their experiment from the drop-down list or set it to Auto-detect.
- *Ionization type*: Ionization methods can affect m/z calculation. Drop-down provides a list of the most popular ionization types.
- *Q1 accuracy*: This is the mass resolution in amu of the first quadrapole.
- Q3 accuracy: This is the mass resolution in amu of the third quadrapole.
- *Filterline*: The drop-down lists different mass ranges and allows the user to process the data in these ranges separately with different parameters. Primarily used for polarity-switching experiments.

File Import

<u>F</u> ile <u>W</u> idgets <u>H</u> elp					
Open Align Peaks Match Options Polly			Auto Detect(+1) Charge 1 . AreaTop	:[m/z]MW / Compound +/- 5.0000	00 : ppm :
🎽 🚔 🔜 🏖 🅦 👾 😑	🙉 <table-cell-columns> 👼 🕹 🔀 👄 🔿 📧 I</table-cell-columns>	тс 📃 🎫 🍫			Ö
Sample name filter					₩ ₩ +†+
					411
					11
					 ≫
					C ⁹
	8 💿 Options				Q
	Instrumentation File Import Peak Detection P	Peak Filtering Isotope Detection EIC (KIC) [BETA] Peak Grouping Group Rank Libraries I	Exter	<u></u>
	 Centroid Scans 				- 0
	Scan Filter Polarity	Not Filtered :	Enable multiprocessing		
	Scan Filter MS Level	Not Filtered :			
	Scan Filter Intensity Minimum Quantile Cutoff				
	Scan Filter: Minimum Intensity	0 ions			
		Use Default Parameters			

- *Centroid Scans*: Centroid acquisition is an acquisition method where only centroid m/z and intensity are stored. Centroid m/z is calculated based on the average m/z value weighted by the intensity and m/z values are assigned based on a calibration file. Users may leave the box unchecked if they have the centroid data or check the box if centroiding has to be done in El-MAVEN.
- *Scan Filter Polarity*: Users may choose to import scans based on the polarity of ions in the scan. Especially helpful in polarity-switching experiments.
- *Scan Filter MS Level*: Users may choose to import only MS1 or MS2 scans. This feature can be used with MS/MS data.
- Scan Filter Minimum Intensity: Sets a minimum threshold for reading in intensity values.
- *Scan Filter Intensity Minimum Quantile Cutoff*: Scans with x% of their intensity values below the threshold will be filtered out during import.
- *Enable Multiprocessing*: In order to reduce the sample load time, El-MAVEN uses multiprocessing. This behavior can be changed by the user.

Peak Detection

<u>F</u> ile <u>W</u> idgets <u>H</u> elp				
Open Align Peaks Match Options Polly		Auto Detect(+1) Charge 1 AreaTop	: [m/z] MW / Compound +/- 5.000000	. ppm :
📁 🖆 🔜 🥙 🌁 🌾 😑	🙉 🛫 👼 🖑 🏊 ⇐ 🔿 🎓 📠 TIC 🔜 🎫 🍫			C
Sample name filter	Options Instrumentation File Import Peak Detection Peak Filtering Isotope Detection E Peak Detection and Grouping Settings Bas EIC Smoothing Algorithm Savitzky-Golay = EIC Smoothing Window 10	seline Calculation prop top x% intensities from chromatogram 80% aseline Smoothing 5 scans	tter	
·				

Peak Grouping and Grouping Settings

- *EIC Smoothing Algorithm*: Smoothing of data points helps in increasing the signal/noise ratio. There are three algorithms provided for EIC smoothing:
 - 1. *Savitzky-Golay*: It preserves the original shape and features of the signal better than most other filters (Learn more).
 - 2. *Gaussian*: It reduces noise by averaging over the neighborhood with the central pixel having higher weight but successfully preserves sharp edges. (Learn more).
 - 3. *Moving Average*: It takes the simple average of all points over time. Signal behavior is not natural. Least preferred method for smoothing (Learn more).
- EIC Smoothing Window: Number of scans used for fitting in the smoothing algorithm can be adjusted here.
- *Maximum Retention Time Difference Between Peaks*: Set a limit to retention time (RT) difference between peaks in a group. Increase the value if alignment fails to center peaks satisfactorily.

Baseline Calculation

- *Drop top x% intensities from chromatogram*: Set the baseline for every peak. Baseline is obtained once x% of the highest intensities in a peak are removed from consideration. Baseline should be set high when there is more noise in the data.
- Baseline Smoothing: Number of scans used for fitting in the smoothing algorithm can be adjusted here.

Peak Filtering

<u>F</u> ile <u>W</u> idgets <u>H</u> elp							
Open Align Peaks Match Options Polly			Auto Detect(+1) Cha	rge 1 :AreaTop	: [m/z] MW / Compound	+/- 5.000000 : ppr	n :
🥣 🚔 🌄 🏖 🌫 🐇 🛸	🙉 🐩 🔂 🗲 🖙 👄	🣂 🔟 TIC 🔜 🗾 🍫					Þ
	© Options Instrumentation File Import Peak I © Isotope Peak Filtering Is same as Parent Peak Filtering Min. Signal Baseline Difference Min. Peak Quality	Detection Peak Filtering Isotope De Peak Filtering 0.00	tection EIC (XIC) [BETA] Peak Group Isotopic Peak Filtering Min. Signal Baseline Difference Min. Peak Quality	0.00			
			Min. reak Quality Parameters				

• *Minimum Signal Baseline Difference*: Set the minimum difference between intensity and baseline to detect any signal as a valid peak.

Isotope Detection

File Widgets Help					
Open Align Peaks Match Options Polly		Auto Detect(+1) Charge 1 A	reaTop : [m/z] N	/W / Compound +/	- 5.000000 . ppm :
🎏 🖆 🔜 🥐 🌁 🌾 😑 🛛 🚳	L 🛠 🛃 🖑 ∑le 🗢 ⇒ 🎓 🔝 TIC 🔜 🖬 🍫				6
Sample name filter					₩ *& [] ††
					+11
					~
					[
					C
80	Options				.
Instru	umentation File Import Peak Detection Peak Filtering Isotope Dete	ection EIC (XIC) [BETA] Peak Grouping Group	Rank Libraries Exter	1	<u></u>
		Filter Isotopic Peaks based on these criteri			م ان
			20 :		
Во	ookmark, peakdetection & file export 🛎 D2 🛎 C13 🛎 N15 🛎 S34		scans :		
lac	otopic widget	Abundance Threshold 1. Maximum %Error to Natural Abundance 10			
130	- 02 - CIS - 1415 - 554	Correct for Natural C13 Isotope Abundan			
	Use Default	Parameters			

Are Samples Labeled?

• *Bookmarks, peak detection, file export*: Select the labeled atoms that should be used in bookmarking, peak detection and export. D2: Deuterium, C13: Labeled carbon, N15: Labeled nitrogen, S34: Labeled sulphur.

- *Isotopic widget*: Select the labeled atoms that should be displayed in the isotopic widget. D2: Deuterium, C13: Labeled carbon, N15: Labeled nitrogen, S34: Labeled sulphur.
- Number of M+n isotopes: Set the maximum number of labeled atoms per ion in the experiment.
- *Abundance Threshold*: Set the minimum threshold for isotopic abundance. Isotopic abundance is the ratio of intensity of isotopic peak over the parent peak.

Filter Isotopic Peaks based on these criteria

- *Minimum Isotope-Parent Correlation*: Set the minimum threshold for isotope-parent peak correlation. This correlation is a measure of how often they appear together.
- *Isotope is within [X] scans of parent:* Set the maximum scan difference between isotopic and parent peaks. This is a measure of how closely they appear together on the retention time scale.
- *Maximum % Error to Natural Abundance*: Set the maximum natural abundance error expected. Natural abundance of an isotope is the expected ratio of amount of isotope over the amount of parent molecule in nature. Error is the difference between observed and natural abundance as a fraction of natural abundance.
- Correct for Natural C13 Isotope Abundance: Check the box to correct for natural C13 abundance.

EIC (XIC) [BETA]

<u>File Widgets H</u> elp					
Open Align Peaks Match Options Polly		Auto Detect(+1) Charge 1 AreaTop	: [m/z] MW / Compound	+/- 5.000000	: ppm :
🎽 🖆 🔜 🧶 ጅ 🌿 😑	🙉 針 👼 🖑 🏊 👄 🔿 📂 📠 TIC 🔜 🎫 🍫				C
Sample name filter	Options Instrumentation File Import Peak Detection Peak Filtering Isotope Detection EIC () EIC Calculation FIC Tyr Max	IIC) [BETA] Peak Grouping Group Rank Libraries Ex	ter - 👔		
	Use Default Parameters				
					8

- *EIC Type*: Select a method to merge EICs over m/z. There are two options:
 - 1. MAX: Merged EIC is created by taking the maximum intensity across the m/z window at a particular scan.
 - 2. *SUM*: Merged EIC is created by taking the sum average of intensities across the m/z window at a particular scan.

Peak Grouping

<u>File</u> <u>W</u> idgets <u>H</u> elp				
Open Align Peaks Match Options Polly		Auto Detect(+1) Charge 1 . AreaTop	: [m/z] MW / Compound +/- 5.000000	: ppm :
💆 🖆 🔜 🥙 🎘 🔆 😑 👘 👘	🙉 🛠 👼 🖑 🏊 👄 🔶 🍅 🔝 TIC 🔜 🎫 🍫			C
Sample name filter	Options Instrumentation File Import Peak Detection Peak Filtering Isotope Detection EIC Peak Grouping Score score = 1.0/((distX*A)+0.01)/ ((distX*B)+0.01)*(C*overlap) where distX = Rt difference of peaks Use Default Parameters	1 1 1	Exter	
				3

- *Peak Grouping Score*: Peaks are assigned a grouping score to determine whether they should be grouped together. There are two formulas for grouping score calculation:
 - 1. score = 1.0/((distX * A) + 0.01)/((distY * B) + 0.01) * (C * overlap)
 - 2. score = 1.0/((distX * A) + 0.01)/((distY * B) + 0.01)

The score depends on the following 3 parameters and their weights:

- *RT difference or DistX*: Difference in retention time between the peaks under comparison. Closer peaks are assigned a higher score.
- *Intensity difference or DistY*: Difference in intensity between peaks under comparison. Smaller difference accounts for a higher score.
- *Overlap*: Fraction of retention time overlap between the peaks under comparison. Greater overlap accounts for a higher score.
 - Uncheck Consider Overlap to calculate grouping score without overlap.
 - Sliders are provided to adjust the weights attached to each of the three parameters.

Group Rank

<u>F</u> ile <u>W</u> idgets <u>H</u> elp					/
Open Align Peaks Match Options Polly			Auto Detect(+1) Charge 1 AreaTop	: [m/z] MW / Compound +/- 5.000000	: ppm :
📁 🚔 🔛 🥙 🌁 🌾 😑	🙉 針 👼 🦨 🏊 🔶 📂 🔝 TIC	E •			C
Sample name filter					₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩ ₩
					411
					\sim
					2
					C° A
	8 💿 Options				e e e e e e e e e e e e e e e e e e e
	Instrumentation File Import Peak Detection Pea Group Rank Formula	k Filtering Isotope Detection EIC (Consider Ret	XIC) [BETA] Peak Grouping Group Rank Libra	aries Exter 🕐	
	Group Rank = ((1.1 - Q)^A) * (1 /(log(I + 1))^B)	Quality Weight -	1.0		0
	Where	Intensity Weight ——	1.0		
	Q = Quality of a group	dRT Weight	1.0		
		Use Default Parameters			
c(-) 10					

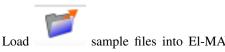
- *Group Rank Formula*: Group rank is one of the parameters for group filtering. There are two formulas below for group rank calculation:
 - 1. Group Rank = $((1.1 Q) \land A) * (1/(\log(I + 1)) \land B)$
 - 2. Group Rank = $((1.1 Q) \land A) * (1/(\log(I + 1)) \land B) * (dRT) \land (2 * C)$

The score depends on the following 3 parameters and their respective weights A, B and C:

- *Q or Group Quality*: Maximum peak quality of a group. Peaks are assigned a quality score by a machine learning algorithm in El-MAVEN. Better quality leads to a higher rank.
- I or Group Intensity: Maximum intensity of a group. Better intensity leads to a higher rank.
- dRT or RT difference: Difference between expected retention time and group mean RT.
 - *Consider Retention Time*: Check the box to use formula (b) for group rank calculation. Formula
 (a) is used by default.
 - Quality Weight: Adjust slider to set weight for group quality in group rank calculation.
 - Intensity Weight: Adjust slider to set weight for group intensity in group rank calculation.
 - *dRT Weight*: Adjust slider to set weight for retention time difference in group rank calculation. The slider is disabled if Consider Retention Time is unchecked.

Sample Upload

Load Sample Files





Load sample files into El-MAVEN and click on *Show Samples Widget* on the widget bar to show/hide the project space. Blanks will not show up in the sample list if the file names start with 'blan' or 'blank'.

File Widgets Help						í.
Open Align Peaks Match Options	Polly			Auto Detect(-1) Charge 1 . AreaTop	: [m/z] MW / Compound +/- 5.000000	: ppm :
		🙉 針 🔂 🗲 🏊 👄 🔿 TIC 🔜	14 🍫			C
Sample name filter				m/z: 482.9589-482.9637		
Sample * Set	Scalin					
		.3.6e+05			ROAL ANO BTOR 15 240M	111
	1.00				6:98:18: 18: 213=728M	T I
091215_120M	1.00					\sim
○ 091215_120i	1.00	3.2e+05				
091215_45i	1.00			•		C ⁹
				T	1.5e465 691213 8M	(A)
	1.00	.2.8e+05				
🗢 091215_5i 🗠	1.00					
■ 091215_3M <						
■ 091215 1M <	1.00	2.4e+05				6
				T		
■ 091215_0i <						
		2e+05				
		1.6e+05				
		1.0+.05				
		1.20+00				
		_Be+04				
		_4e+04				
				AND AND		
ia		12800000	14.31	14.81 15.31	15.81 16.31	

Load sample files into El-MAVEN and click on *Show Samples Widget* on the widget bar to show/hide the project space. Blanks will not show up in the sample list if the file names start with 'blan' or 'blank'.

There are three columns in the project space:

- *Sample*: This column has the sample name and the random color assigned to the sample. Double-click the sample name to change the color.
- Set: The set column holds the cohort name for every sample. For example: subjects and controls.
- *Scaling*: This column holds the normalization constant for every sample. For example, all intensities in a sample will be halved if the constant is two. This is done to normalize data if sample volumes are different.

Sample Space Menu

Load Project: Sample files can be loaded here.



Load Meta: Users may upload a meta file with sample and set names in a comma separated file (.csv) or double-click to enter text. Meta file template is shown below:

	Α	В
1	Sample	Set
2	091215_5M	Test
3	091215_5i	Control
4	091215_45M	Test
5	091215_45i	Control
6	091215_3M	Test
7	091215_1M	Test
8		

Save Project as: Current state of El-MAVEN can be saved in a .mzroll file for future use. All the settings, EICs and peak tables are stored in the file and may be reloaded at any point in the future.



Change Sample Color: Sample colors can be changed by either clicking on this menu button or double-clicking the sample name. Users can pick a color of their choice to represent their samples.



Remove Samples: Apart from deselecting samples, users also have the option to remove samples from the project space. The sample files will not be deleted, only removed from El-MAVEN's project space.



Show/Hide Selected Samples: Samples can be selected/deselected in batches. This is especially helpful when dealing with large datasets as the EIC window gets increasingly noisy with more samples.

Mark Sample as Blank: Users can select sample files and set them as blanks as depicted below. Clicking the button again will reverse the move.

File Widgets Help	<u>မှ</u> 🗶 🚺							5 (3)		- UK - 1
Open Align Peaks N	latch Options Pol	ly				Auto Detect(-1) Charge 1 . A	Arealop	: [m/z] MW / Compound	+/- 5.000000	: ppm :
📁 🥔 🔜 🖉 🛣	⅍ ━	🙉 📌 👼 🧧	🗲 🏊 🔶 🃂 🚺	ti TIC 📃 🎫 🍫						C
Sample name filter	Set As a Blan	k Sample				v/z: 482.9589-482.9637				₩6 11 †1
Sample	Set 9	icalin								a da
 exampleProject 	_	3.6e+05						Roak Area T9815 240M		1+1
Ø 091215_240 Ø 091215_240		.00						6:3818181818181818		TI+
S 091215_240 S 091215_120	I <mark>M</mark>	.00						16-05 6 1215 45M		\sim
Ø 091215_120		.00 3.2e+05								۵
Ø 091215_45N Ø 091215_45i		.00						S:48-03 8 2 3 3 M		C
091215_15N	<mark>1</mark>	.00				T		768265 881218 8M		ě
091215_15i 091215_5M		.00 .00 2.8e+05								
091215_5i	د 1	.00								[
■ 091215_3M ■ 091215_3i		.00								
■ 091215_51 ■ 091215_1M		.00 2.4e+05								1
■ 091215_1i	د 1	.00								
■ 091215_0M ■ 091215_0i		.00								
	-	2e+05								
		20+05								
		_1.6e+05				1				
		_1.2e+05								
		8e+04								
		_4e+04								
						and all				
		290000		20a0 	14.31	14.81	15.31	15.81	16.31	
Unknown compound Cl	liphoard set to									3

Compound Database

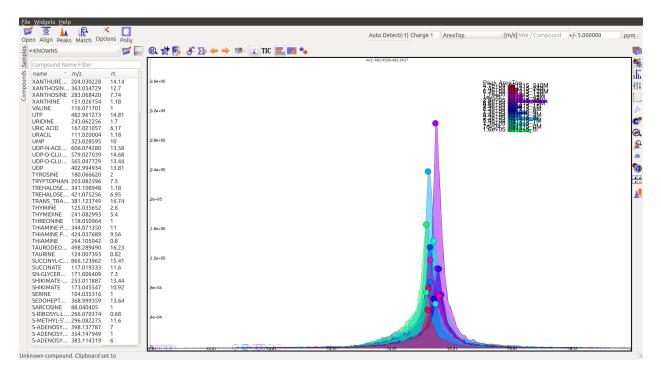
Load Reference File

	A	В	С	D	E
1	compound	formula	id	rt	category
2	cholesteryl sulfate	C27H46O4S	HMDB00653	17.25	
3	Deoxycholic acid	C26H43NO5	C05464	16.79	
4	trans_trans-farnesyl diphosphate	C15H28O7P2	C00448	16.74	
5	Cholic acid	C24H40O5	HMDB00619	16.69	
6	Geranyl-PP	C10H20O7P2	C00341	16.46	
7	Taurodeoxycholic acid	C26H45NO6S	C05463	16.23	
8	lipoate	C8H14O2S2	C00725	15.97	
9	3-hydroxy-3-methylglutaryl-CoA-nega	C27H44N7O20P3S	C00356	15.72	
10	butyryl-CoA	C25H42N7O17P3S	C00630	15.72	

Reference file contains a list of metabolites and their properties that are used for peak detection. This is a comma separated (.csv) or tab separated (.tab) file with compound name, id, formula, mass, expected retention time and category. It is preferable but not necessary to have retention time information in the reference file but either mass or formula is required. In case both mass and formula are provided, formula will be used to calculate the m/z. Click on the *Show Compounds Widget* on the widget toolbar to view the compounds panel. Users may upload a new reference file or use any of the default files loaded on start-up.

File Widgets Help Image: State of the		Auto Detect(-1) Charge 1 . AreaTop	: [m/z] MW / Compound +/- 5.000000	: ppm :
	🙉 🛠 👼 🚜 듲 ⇒ 📂 😰 TIC 📃 🎫 🍫			Þ
Compound Name Filter		m/z: 482.9589-482.9637		*
eg name m/z rt NANTHUSIN 204.030228 14.14 XANTHOSIN 363.034729 12.7 XANTHOSINE 283.068420 7.74	3.6e+05		Deale Area Terr	ահ
ANTHORE 204.030228 14.14 XANTHOSIN 363.034729 12.7	5.0000		4.78184 84 315_348M	+11
5 XANTHOSINE 283.068420 7.74			6:38103 881213=128M	Γ. /
Ö XANTHINE 151.026154 1.18				
VALINE 116.071701 1	3.2e+05		8:28+84 88 215-32	2
UTP 482.961273 14.81			\$:48183 88 5 5 SM	-0
URIDINE 243.062256 1.7		•	7.38485 881318 -1M	C°
URIC ACID 167.021057 6.17 URACIL 111.020004 1.18			1.6e+05 091215 01"	۵,
UMP 323.028595 10	2.8e+05			
UDP-N-ACE 606.074280 13.58				R
UDP-D-GLU 579.027039 14.68				۵
UDP-D-GLU 565.047729 13.44				
UDP 402.994934 13.81	2.4e+05			0
TYROSINE 180.066620 2				
TRYPTOPHAN 203.082596 7.5 TREHALOSE 341.108948 1.18				
TREHALOSE 341.108948 1.18 TREHALOSE 421.075256 6.95				A.
TRANS TRA 381.123749 16.74	2e+05			
THYMINE 125.035652 2.6				
THYMIDINE 241.082993 5.4				
THREONINE 118.050964 1				
THIAMINE-P 344.071350 11	_1.6e+05	T		
THIAMINE P 424.037689 9.56				
THIAMINE 264.105042 0.8 TAURODEO 498.289490 16.23				
TAURINE 124.007393 0.82				
SUCCINYL-C 866.123962 15.41	_1.2e+05			
SUCCINATE 117.019333 11.6				
SN-GLYCER 171.006409 7.3				
SHIKIMATE 253.011887 13.44				
SHIKIMATE 173.045547 10.92	_Be+04			
SERINE 104.035316 1 SEDOHEPT 368.999359 13.64				
SARCOSINE 88.040405 1				
S-RIBOSYL-L 266.070374 0.68				
S-METHYL-5' 296.082275 11.6	4e+04			
S-ADENOSY 398.137787 7		A AL		
S-ADENOSY 354.147949 1		A A A A		
S-ADENOSY 383.114319 6		and the		
Contraction of the second sectors	1280 000 13.31 O @ (BD)(3.81) 14.31	14.81 15.31	15.81 16.31	

EIC



An Extracted Ion Chromatogram is a graph of Intensity vs. retention time for a certain m/z range. EIC window displays the EIC for every group/compound selected or m/z range provided. The group name and/or the m/z range is displayed at the top. Following are the different menu options on top of the EIC window:



Zoom Out: The EIC is initially zoomed-in to display the region near the expected retention time of a group. This button will zoom out and display the whole retention time range for the selected m/z range. Users may zoom in to a region by right dragging the mouse over it. Left-dragging will zoom out.

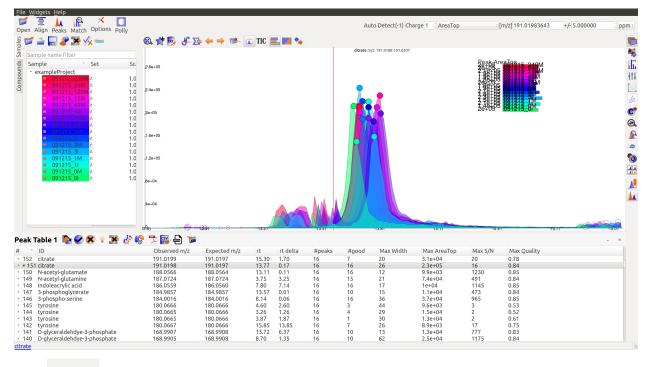


.

Copy Group Information to Clipboard: On clicking this button, group information is copied to the clipboard with every row representing a different sample.

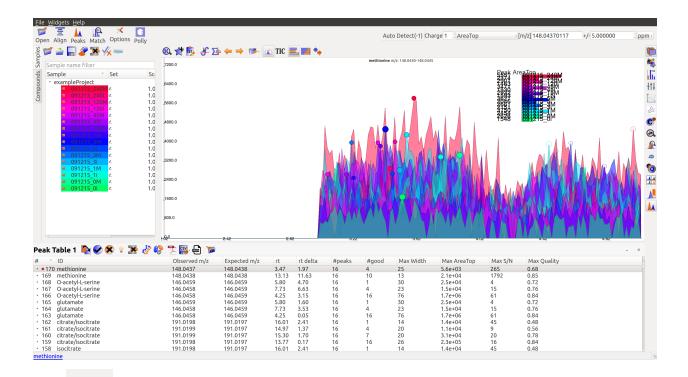


Bookmark as Good Group: Users can manually curate a group as 'good' and store it in the bookmark table using this button. (Manual curation of groups has been covered here)





Bookmark as Bad Group: User can manually curate a group as 'bad' and store it in the bookmark table using this button. (Manual curation of groups has been covered here)



History Back: EIC window display history is recorded. Clicking this button will display the previous state of the window.

-

•

History Forward: EIC window display history is recorded. Clicking this button will display the next state of the window, if available.



Save EIC Image to PDF File: Saves the current EIC window display in a PDF file.



Copy EIC Image to Clipboard: Current EIC window display is copied to clipboard.



Print EIC: Current EIC window display can be directly printed out.



Auto Zoom: Auto Zoom is selected by default. It zooms-in and centers the EIC to the expected RT. The expected retention time is depicted as a dashed red line.

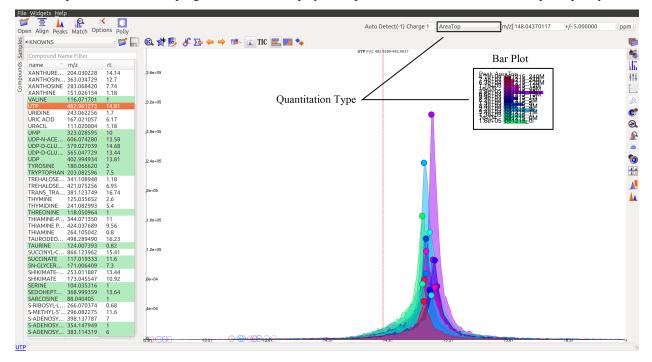
P		. (A)	< 💽				Auto Detect(-1) Charge 1	AreaTop	: [m/z] 148.04370117	+/- 5.000000	: ppm :
Ор	en Align Peak	s Match Opt	ions Polly				rideo bececi (i) enorge i	. / incurop	- [iii/c] i loid ibi di li	.7 5.000000	. ppm.
les	■KNOWNS		: 💅 💭	🙉 針 👼 🕂 🅦	🔶 🏓 🥦 🔝 TIC 🔜 🜉	•					r i
Samp							UTP m/z: 482.9589-482.9637				
	Compound Na										* 1
spunodmo;	name *	111/2	rt								11
no	XANTHURE		14.14	_3.6e+05					2924 01 200 200 200 200 200 200 200 200 200		+++
đ	XANTHOSIN XANTHOSINE		12.7					(5:38+84 631 318_728M		111
S		151.026154	1.18						286455 181213 43M		\sim
	VALINE	116.071701	1	_3.2e+05							
	UTP	482.961273	14.81	5.20+00							
	URIDINE	243.062256	1.7						2:38+88 881218 =1M		C ⁹
	URIC ACID	167.021057 111.020004	6.17 1.18		Expec	ted RT 🗔			1.6e+05 091215 01		
	UMP	323.028595	10	_2.8e+05	LAPCE						~
	UDP-N-ACE		13.58								R
	UDP-D-GLU		14.68								
	UDP-D-GLU		13.44								
	UDP TYROSINE	402.994934 180.066620	13.81 2	_2.4e+05			•				0
	TRYPTOPHAN		7.5				1				
	TREHALOSE		1.18				A				
	TREHALOSE		6.95				· · · · ·				
	TRANS_TRA		16.74	_2e+05							4.
	THYMINE	125.035652	2.6 5.4								
	THREONINE		1								
	THIAMINE-P		11	_1.6e+05							
	THIAMINE P		9.56								
	THIAMINE	264.105042	0.8					2			
	TAURODEO TAURINE	498.289490	16.23 0.82								
	SUCCINYL-C		15.41	_1.2e+05							
	SUCCINATE	117.019333	11.6								
	SN-GLYCER		7.3								
	SHIKIMATE		13.44								
	SHIKIMATE	173.045547 104.035316	10.92	_8e+04							
	SEDOHEPT		13.64								
	SARCOSINE		1								
	S-RIBOSYL-L		0.68	.4e+04							
	S-METHYL-5'		11.6								
	S-ADENOSY S-ADENOSY		7				JAP P				
	S-ADENOSY		6				and a set				
	C. C					19.41	14.01		Inat	10.01	
				the two successions		/ 1001					

TIC

Show TICs: Displays the Total Ion Current. TIC is the sum of all intensities in a scan.



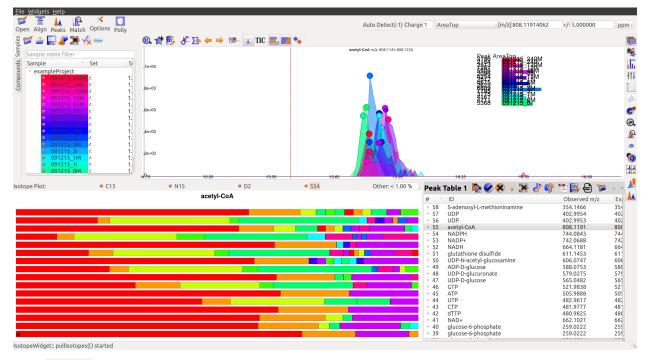
Show Bar Plot: Displays the peak intensity for a group in every sample. Intensity can be calculated by various methods known as quantitation types in El-MAVEN. Users can change the quantitation type from the drop-down list on the top right or choose to display other parameters like retention time and peak quality.





•

Show Isotope Plot: Displays the isotope plot for a group. Each bar in the plot represents the relative percentage of different isotopic species for the selected group in a sample.





Show Box Plot: Displays the boxplot for a group. The box plot shows the spread of intensities in the group and where each peak lies in relation to the median. Median of the intensities is the vertical line between the boxes.

Apart from the top menu, there are other features in the EIC window. Right-click anywhere in the window and go to Options.

Align Peaks Match Op	K 🚺			Auto Detect(-1) Cha	rge 1 . AreaTop	: [m/z] 808.11914062	+/- 5.000000	: pp
NOWNS	: 📁 📊	🙉 📌 👼 子 🏊	🔶 🔿 🌮 🔝 TIC 🔜 🎫 🍫					
ompound Name Filter				UTP m/z: 482.9589-482.9637				
ame * m/z	rt h							
ANTHURE 204.030228	14.14	3.6e+05						
ANTHORE 204.030228 ANTHOSIN 363.034729	12.7	5.00403				4.72+04 001215 240M		
ANTHOSINE 283.068420	7.74					6:38184 881 2185538M		
ANTHINE 151.026154	1.18					296765 881315 45M		
ALINE 116.071701	1					9:98+84 89-2 13-13		
TP 482.961273	14.81	.3.2e+05				8:38±83 88-3 +8=8M		
RIDINE 243.062256	1.7				•	5:28182 831213 TM		
RIC ACID 167.021057	6.17				- -	7820782 89 213 AM		
RACIL 111.020004	1.18					1.66+05 091215-01		
MP 323.028595		2.8e+05						
DP-N-ACE 606.074280	13.58							
DP-D-GLU 579.027039	14.68							
DP-D-GLU 565.047729	13.44							
DP 402.994934	13.81	2.4e+05						
(ROSINE 180.066620	2	.2.46+05						
RYPTOPHAN 203.082596	7.5		Peak Grouping Option	ns				
REHALOSE 341.108948	1.18		Add a Note					
REHALOSE 421.075256	6.95		Recalculate EICs					
RANS TRA 381.123749	16.74	2e+05	Copy EIC(s) to Clipboa	rd				
TYMINE 125.035652	2.6		Options	 Show Peaks 				
YMIDINE 241.082993	5.4		Options					
HREONINE 118.050964	1			 Group Peaks Automatically 				
HAMINE-P 344.071350	11	1.6e+05		 Show EIC 				
HAMINE P 424.037689	9.56	0.00100		Show Spline				
HAMINE 264.105042	0.8			Show Cubic Spline				
URODEO 498.289490	16.23			 Show Baseline 				
AURINE 124.007393	0.82			Show TIC				
JCCINYL-C 866.123962	15.41	1.2e+05		 Show Bar Plot 				
JCCINATE 117.019333	11.6			Show BIC				
N-GLYCER 171.006409	7.3			Show Merged EIC				
HIKIMATE 253.011887	13.44			Show EICs as Lines				
HKIMATE 173.045547	10.92	8e+04						
RINE 104.035316	1			Show Box Plot				
DOHEPT 368.999359	13.64			Show MS2 Events				
ARCOSINE 88.040405	1							
RIBOSYL-L 266.070374	0.68	4e+04						
METHYL-5' 296.082275	11.6	140104						
ADENOSY 398.137787	7			- A				
ADENOSY 354.147949	1			of all a				
ADENOSY 383.114319	6			and the second				
		Las COM				Star in the second s		

Some of the important options are:

- *Show Peaks*: Peaks are marked by the colored circles that represent the quality score of the peak. Bigger the circle, better the peak quality. This option allows the user to show/hide the peak quality score.
- *Group Peaks Automatically*: Peak grouping happens automatically when grouping parameters are changed. To prevent automatic grouping, user can uncheck this option.
- Show Baseline: Hide/Show the baseline for every peak. (Read more about baseline here).
- *Show Merged EIC*: Merged EIC is the sum average of EICs across samples. It smoothens the data and helps in grouping peaks.
- *Show EIC as Lines*: In case of large number of samples, it can get difficult to look at short individual peaks as they are obscured by larger peaks. Showing EIC as lines cleans up the display window and allows the user to look at small peaks.

Mass Spectra

Mass Spectra Widget displays each peak, its mass, and intensity for a scan. As the widget shows all detected masses in a scan, the ppm window for the EIC and consequently grouping can be adjusted accordingly. This feature is especially useful for MS/MS data and isotopic detection.

Fil	e <u>W</u> idgets <u>H</u> elp								
0	📁 \Xi 🛕 🔐 🏓						Auto Detect(-1) Charge 1 . AreaTop	: [m/z] 808.11914062	+/- 5.000000 i ppm :
ples	KNOWNS Compound Name Filter name m/z XANTHURE 204.030228 XANTHOSIN 363.034729 XANTHOSINE 283.068420 XANTHINE 151.026154	: 📁 🔚	🙉 쑭 👼	\$ ∑⊨ 👄 ⇒	🃂 主 TIC 📃	14 🍫			6
Is San	Compound Name Filter	rt	.3.6e+05				UTP m/z: 482.9589-482.9637	2:48144 03 315-340M	***
bound	XANTHURE 204.030228 XANTHOSIN 363.034729	14.14 12.7	_3.2e+05				•	6.38+84 88 213 128M 8.884184 88 213 13M	1111 111
Com	XANTHOSINE 283.068420 XANTHINE 151.026154	7.74	.2.8e+05						
	UTP 482.961273	1 14.81	,2.4e+05					1.5e+05 891218 0M	<u>ام</u>
	URIDINE 243.062256 URIC ACID 167.021057 URACIL 111.020004	1.7 6.17 1.18	1.6e+05						[⊻
	UMP 323.028595 UDP-N-ACE 606.074280	10 13.58	_1.2e+05				8		<u>_</u>
	UDP-D-GLU 579.027039 UDP-D-GLU 565.047729 UDP 402.994934	14.68 13.44 13.81	_8e+04						
	TYROSINE 180.066620 TRYPTOPHAN 203.082596	2	_4e+04						1
SDE	ectra		129 80 @@@	13.31	13:810	14.31	14.81 10.31	10.61	
3.6e					091215_45i sca	n# 998 rt: 15.21 ms:1 ion:- p	orodMz: 482.961		
3.2e	+05								
2.8e	+05			482.	961395 ^{z=1}				
2.4e	+05								
2e+I	05								
1.6e	+05								
.1.2e	+05								
ße+I	04								
4e+I	04					483.964752			
Bo.s		481.977600	482.36		483.06	483.76	484.965759 485.16	485.86485.972137	486.56

Alignment

Prolonged use of the LC column can lead to a drift in retention time across samples. Alignment shifts the peak RTs in every sample to correct for this drift and brings the peaks closer to median retention time of the group.



Click on the *Align* button

and adjust the settings.

		Alignment Options e 1 AreaTop : [m/z] 808.11914062 +/-5.000000	lane.
Open Align Peaks Match Options Polly		Group Selection Criteria for use in Alignment	: ppm :
š rKNOWNS : 💕 🚽	🙉 😤 👼 🕹 🏊 👄 🔿	Group must contain at least [X] good peaks 1 peaks	0
Sing *KNOWNS : 20 - 20 - 20 - 20 - 20 - 20 - 20 - 20		Limit total number of groups in alignment to 1000 groups	1
ଓ Compound Name Filter ତ name ^ m/z rt		Peak Grouping Window 10 scans	
P0 name m/z rt XANTHURE 204.030228 14.14 XANTHOSIN 353.034729 12.7 XANTHOSIN 283.068420 7.74	3.6e+05	Alignment w.r.t Expected Rt	
ANTHORE 204.030228 14.14 XANTHOSIN 363.034729 12.7	1.0000		+†∔
5 XANTHOSINE 283.068420 7.74		Peak Selection	To 1
- XANTHINE 151.026154 1.18		Minumum Peak Intensity 5000.00	
VALINE 116.071701 1	_3.2e+05	Minimum Peak S/N Ratio 2	۵
UTP 482.961273 14.81 URIDINE 243.062256 1.7		Minimum Peak Width 3 scans	C
URIC ACID 167.021057 6.17			U
URACIL 111.020004 1.18		Peak Detection Algorithm Compound Database Search	
UMP 323.028595 10	_2.8e+05	Select Database KNOWNS	R
UDP-N-ACE 606.074280 13.58			
UDP-D-GLU 579.027039 14.68			
UDP-D-GLU 565.047729 13.44		Alignment Algorithm Poly Fit :	
UDP 402.994934 13.81 TYROSINE 180.066620 2	.2.4e+05		0
TRYPTOPHAN 203.082596 7.5		Maximum Number of Itterations 10	
TREHALOSE 341.108948 1.18		Polynomial Degree 5	
TREHALOSE 421.075256 6.95			
TRANS_TRA 381.123749 16.74	_2e+05		1.
THYMINE 125.035652 2.6			-
THYMIDINE 241.082993 5.4 THREONINE 118.050964 1			
THREONINE 118.050964 1 THIAMINE-P 344.071350 11	_1.6e+05	D	
THIAMINE P 424.037689 9.56	1.6e+05		
THIAMINE 264.105042 0.8			
TAURODEO 498.289490 16.23			
TAURINE 124.007393 0.82	1.2e+05		
SUCCINYL-C 866.123962 15.41	1.20403		
SUCCINATE 117.019333 11.6 SN-GLYCER 171.006409 7.3			
SHIKIMATE 253.011887 13.44			
SHIKIMATE 173.045547 10.92	.8e+04		
SERINE 104.035316 1			
SEDOHEPT 368.999359 13.64			
SARCOSINE 88.040405 1			
S-RIBOSYL-L 266.070374 0.68	.4e+04	Status Undo Aligment	
S-METHYL-5' 296.082275 11.6			
S-ADENOSY 398.137787 7 S-ADENOSY 354.147949 1	1	0% Alian Cancel	
S-ADENOSY 334.147949 1 S-ADENOSY 383.114319 6			
······································	12/81) 13.31	10.21 10.21 10.31	
UTP			13

The first panel in Alignment options is for Group Selection criteria. 'Group' here refers to a set of peaks across samples that is annotated as a particular ion.

- *Group must contain at least [X] good peaks*: The value of x is set to filter out groups that do not have at least x good peaks from the alignment process. As there is only one peak per sample for a group, this value should not exceed the number of samples in your project. This option allows the users to discard groups with very few good peaks under the assumption that those could be stray peaks.
- *Limit total number of groups in alignment to*: Users can change the number of groups being used for alignment in case there are too many groups detected after the peak detection process.
- *Peak Grouping Window*: This value controls the number of scans required to get the most accurate peaks. Enter a high number if the reproducibility is low to ensure best results.

The next panel is for Peak Selection settings:

- *Minimum Peak Intensity*: The intensity value can be adjusted to only look at high or low intensity peaks in case you have prior information about the concentration of metabolite users are looking for.
- *Minimum peak S/N ratio*: This is the minimum signal to noise ratio of your experiment. Increase the value if you see too much noise in the data.
- Minimum Peak Width: This is the least number of scans to be considered to evaluate the width of any peak.
- *Peak Detection Algorithm*: Select the *Compound Database Search* algorithm and then choose an appropriate database from the next drop-down menu.

The Alignment Algorithm panel provides the following options:

- *Alignment Algorithm*: There are three alignment algorithms available in El-MAVEN: Obi-Warp, Poly fit and Loess fit. Loess fit has been released as a beta feature for now.
- *Maximum number of Iterations*: This parameter is only required for Poly fit algorithm. Enter the number of times El-MAVEN should fit a model to the data in order to align it.
- *Polynomial Degree*: This is the degree of the non-linear model we are trying to fit. Recommended settings are entered by default.

Click on *Align* at the bottom.

Alignment Visualizations

El-MAVEN provides three visualizations for alignment analysis.

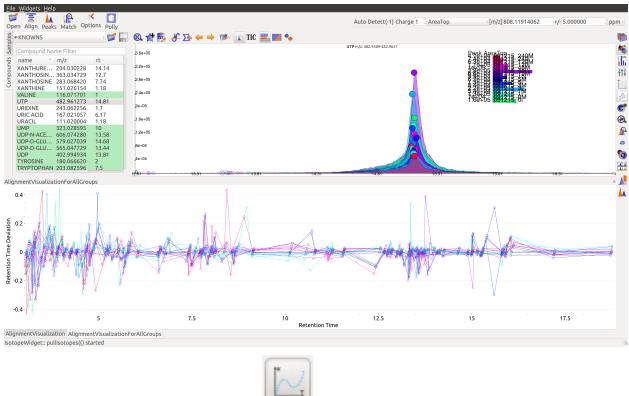


• *Show Alignment Visualization*: Click on **Show Alignment Visualization**: Click on any grouped peak to look at its delta RT vs RT graph as shown.

in the widget bar for this visualization.

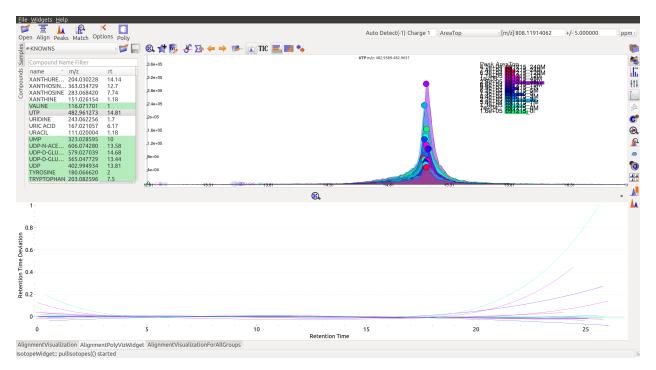
Fi	le <u>W</u> idgets <u>H</u> elp												
	🔰 \Xi 🛕 🔐							Auto Detect(-1) Cha	arge 1 . AreaTop	: [m/z] 808.11914062	+/- 5.000000	: ppn	m :
s	pen Align Peaks Match 9		~ ··· •	10 - 1									nes.
nple	KNOWNS Compound Name Filter name m/z XANTHURE 204.03022 XANTHOSIN 363.03472 XANTHOSINE 283.06842 XANTHINE 151.02615	: 📁 📩	🔍 🏋 📂	, 🕹 🏹 🔶 🔿	🕨 🏊 🚺	C 🔜 🖬 '	•						Þ
San	Compound Name Filter		,3.6e+05					UTP m/z: 482.9589-482.9637		Peak AreaTop15 240M			₩
spu	name * m/z	rt								6:48788 18 1218_238M			1
INOC	XANTHURE 204.03022 XANTHOSIN 363.03472		_3.2e+05						•				4t1
mo	XANTHOSINE 283.06842	0 7.74	.2.8e+05						1				L.
0	XANTHINE 151.02615 VALINE 116.07170	4 1.18	2.4e+05										[∠. ≽
	UTP 482.96127									188465 891213 8M			2
	URIDINE 243.06225	6 1.7	_2e+05										C° Q
	URIC ACID 167.02105 URACIL 111.02000		_1.6e+05										۵,
	UMP 323.02859	5 10	1.2e+05										P
	UDP-N-ACE 606.07428 UDP-D-GLU 579.02703												
	UDP-D-GLU 565.04772	9 13.44	_Be+04										
	UDP 402.99493 TYROSINE 180.06662		_4e+04										0
	TRYPTOPHAN 203.08259		-A		1777 D 00 00			and the					
Ali	gnmentVisualization		12.81 0 000	13.31		3.81	14.31	14.61	10.31	10.81	10.31		
	R-squared = -0.15303			1			_					1	Ĩ
	Compound RT = 14.81000												-
ŝ													
Samples													
Sar													
		14.5				15	Retention Time		15.5		16		
Δli	ignmentVisualization Align	mentVisualization	ForAllGroups				Recención Time						
	topeWidget:: pullisotopes((. ensurer oups										8

• Show Alignment Visualization (For All Groups): Click on



• Show Alignment Polynomial Fit: Click on

in the widget bar for Poly fit alignment.



The above graphs give a clear indication of how aligned/misaligned the peaks are. Users may run alignment again with different parameters if required (or with a different algorithm).

Peak Detection

Peak detection algorithm pulls the EICs, detects peaks and performs grouping and filtering based on parameters controlled by the users. The algorithm groups identical peaks across samples and calculates the quality score by a machine

learning algorithm. Click on the Peaks icon

on the top to open the settings dialog.

There are 3 tabs for setting Peak Detection parameters:

1. Feature Detection Selection

KNOWNS	ks Match Opt		💿 🚜	👼 🖑 ∑‡ ← → 🃂 💽 TIC 📰 📷	• _					1
					' .	UTP	n/z: 482.9589-482.9637			
Compound Na	ame Filter									1
name *	m/z	rt								
XANTHURE	204.030228	14.14	_3.6e+05							
XANTHOSIN		12.7		😣 💿 Peak Detection						
XANTHOSINE		7.74		The second secon						
XANTHINE	151.026154	1.18		Feature Detection Selection Group Filtering Me	ethod Summai	ry				
VALINE UTP	116.071701	1	.3.2e+05	• Automated Feature Detection: Find peaks	by slicing m/	z and retentio	n time space			
URIDINE	482.961273 243.062256	14.81		Mass Domain Resolution (ppm) 20.000		m/z Range	0.00	: 100000000.00 ;		
URIC ACID	167.021057	6.17								
URACIL	111.020004	1.18		Time Domain Resolution (scans) 10 scans	. Limit	Time Range	0.00	: 100000000.00 :		
UMP	323.028595	10	2.8e+05	Auto Detect And Ignore Isotopes	Limit	Intensity Range	50000.00	9999999999.00		
UDP-N-ACE		13.58								
UDP-D-GLU	579.027039	14.68		Compound Database Search: Limit slices to	set of know	n m/z and rete	ntion time valu	es		
UDP-D-GLU	565.047729	13.44		Select Database	1	KNOWNS		:		
UDP	402.994934	13.81	2.4e+05	EIC Extraction Window +/- PPM		5.00 ppm				
TYROSINE	180.066620	2		EIC EXCIACCION WINDOW +/- PPM		5.00 ppm				
TRYPTOPHAN		7.5		Match Retention Time (+/-)		1.00 min				
TREHALOSE	. 341.108948	1.18		Limit Number of Reported Groups per Compo	und I	10 best				
TREHALOSE		6.95	2e+05	clinic Number of Reported Groups per compo	Juliu	To best				
TRANS_TRA THYMINE	125.035652	16.74 2.6		Match Fragmenation		Report Isotop	ic Dooks			
	241.082993	5.4		Precursor Match Tollerance 0.00		Report isotop	ic reaks			
THREONINE		1								
THIAMINE-P		11	1.6e+05	Fragment Peak Matching Tollerance 0.00			Isotope Detecti	on Options		
THIAMINE P		9.56	1.00+00	Minumum Match Score 0.00						
THIAMINE	264.105042	0.8		Million Macch Score 0.00						
TAURODEO	498.289490	16.23								
TAURINE	124.007393	0.82								
SUCCINYL-C		15.41	_1.2e+05	Done						
SUCCINATE		11.6		10	00%			Find Peaks Cancel		
SN-GLYCER		7.3								
SHIKIMATE		13.44			Use Default P	arameters				
SHIKIMATE	173.045547	10.92	_8e+04		obe berooter	arameters				
SERINE	104.035316	1								
SEDOHEPT		13.64								
SARCOSINE		1								
S-RIBOSYL-L.		0.68	4e+04							
S-METHYL-5'		11.6								
S-ADENOSY S-ADENOSY		1					A. P			
	. 354.147949	6					and all all all			

The Feature Detection Selection panel has the following parameters:

- Automated Feature Detection: This is one of the two strategies for finding peaks. Automated search creates thousands of mass slices across the whole m/z and retention time space to find all peaks present in the sample. This strategy is used when looking for new/unknown metabolites in the samples.
 - Mass Domain Resolution: This value defines the m/z range of every mass slice in parts per million
 - *Time Domain Resolution*: This value defines the scan range (or retention time range) of every mass slice
 - *Limit Mass Range*: User can limit the automated search to a range of m/z according to their requirements
 - *Limit Time Range*: User can limit the automated search to a retention time range according to their requirements
- *Compound Database Search*: Database search is used to search for compounds listed in the reference file using their m/z information. For better accuracy, retention time information can also be used for the search.
 - Select Database: Select a desired reference file for the search from the drop-down list
 - *EIC Extraction Window*: Provide a ppm buffer range to all compound masses. A larger window is useful for processing low resolution data. The window should be smaller for high resolution data to reduce noise.
 - Match Retention Time: Enable/disable use of retention time information along with m/z to perform database search. Compounds can have different retention times in every experiment, therefore this option should only be checked if the reference file is specific to the experiment and the sampled used. Enter the time buffer in the accompanying box.
 - *Limit Number of Reported Groups Per Compound*: Multiple groups can be annotated as the same compound, especially when retention time is not taken into consideration for the search.

Users can set the value to only report X best groups according to their rank. The group rank formula will be discussed later in the tutorial.

- Match Fragmentation: This panel is activated for MS/MS data.
- Report Isotopic Peaks: Check this box to find and report isotopic peaks for labeled data.

To perform peak detection with reference, check the box next to *Compound Database Search* and choose the appropriate database. The *EIC Extraction Window* should be set according to the instrument's resolving power. Select the *Match Retention Time* option if you wish to search for compounds using both the m/z ratio and retention time value given in the database. In case of a generic database, searching by retention time is not recommended.

2. Group Filtering

NOWNS	s Match Opt		o				
		🔽 🔚		▶ J. 2. + > 1. IC = 1. + +			
ompound Na	me Filter			OTHING POLICION			
ame î	m/z	rt					
ANTHURE	204.030228	14.14	.3.6e+05		_		
ANTHOSIN		12.7		😣 🐵 Peak Detection			
ANTHOSINE		7.74					
	151.026154	1.18		Feature Detection Selection Group Filtering Method Summary			
	116.071701	1	3.2e+05	Group Filtering			
	482.961273	14.81		Minimum Peak Intensity 5,000.00 AreaTop			
	243.062256 167.021057	1.7					
	111.020004	6.17 1.18		At least 🛛 👘 👘 👘 👘 At least 🖓			
	323.028595	10	2.8e+05	Minimum Quality 0.50			
	606.074280	13.58					
	579.027039	14.68		At least 🖓 👘 1 peak above minimum quality			
DP-D-GLU		13.44					
	402.994934	13.81					
	180.066620	2	.2.4e+05	Minimum Signal/Blank Ratio 2.00			
	203.082596	7.5		At least O 1 peak above minimum signal/bla	nk ratio		
REHALOSE		1.18					
REHALOSE		6.95		Min. Signal/BaseLine Ratio 2			
RANS TRA		16.74	_2e+05	At least 🛛 🚽 🖉 At least At	eline ratio		
	125.035652	2.6					
HYMIDINE	241.082993	5.4		Minimum Peak Width 3 scans			
HREONINE		1					
HIAMINE-P		11	1.6e+05	Min. Good Peak / Group 1 peaks			
HIAMINE P		9.56		Peak Classifier Model File x2/Projects/ElMaven/bin/default.model Load Model			
	264.105042	0.8					
	498.289490	16.23					
	124.007393	0.82	1.2e+05				
	866.123962	15.41	1.20100	Done			
	117.019333	11.6		100% Find Pea	ks Cancel		
	171.006409	7.3					
	253.011887	13.44		Use Default Parameters			
	173.045547 104.035316	10.92	_8e+04				
	368.999359	13.64					
ARCOSINE		13.04					
RIBOSYL-L		0.68					
	296.082275	11.6	_4e+04				
	398.137787	7					
	354.147949	1					
ADENOSY		6					

After grouping is done, groups that do not fulfill the criteria shown above are filtered out.

- *Minimum Peak Intensity*: Groups with no peak intensities above this threshold are filtered out. The drop-down list beside the input box defines how intensity is calculated. Different methods of intensity calculation are known as quantitation types. The slider below can be adjusted to change the minimum percentage of peaks per group that must pass the threshold (minimum number of peaks is 1).
- *Minimum Quality*: Quality of peaks is calculated using a machine learning algorithm. Groups with no peak qualities above this threshold are filtered out. The slider below can be adjusted to change the minimum percentage of peaks per group that must pass the threshold (minimum number of peaks is 1).
- *Minimum Signal/Blank Ratio*: Signal/Blank ratio is the ratio of peak intensity over maximum intensity observed in blanks. Groups with no peaks above this threshold are filtered out. The slider can be adjusted to change the minimum percentage of peaks per group that must pass the threshold (minimum number of peaks is 1). This helps in filtering out peaks that are also present in blanks.
- *Minimum Signal/Baseline Ratio*: Signal/Baseline ratio is the ratio of peak intensity over baseline value for that peak. Baseline calculation is used to filter out noise in the signal and will be discussed later in the tutorial. The slider can be adjusted to change the minimum percentage of peaks per group that must pass the threshold (minimum number of peaks is 1).

- *Minimum Peak Width*: Peak width is equal to the number of scans that a peak is spread over. Groups with no peak widths above this threshold are filtered out. Spurious signals can be filtered out using this option.
- *Peak Classifier Model File*: This is the default model that is used by the machine learning algorithm for classifying peaks according to their quality.

Change the settings according to the data and click on *Find Peaks* to run peak detection. For beginners, performing peak detection with default values at first is recommended. Users may then adjust the settings depending on their results.

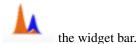
3. Method Summary

KNOWNS		. 🛃 🥅	🙆 🚜	👼 🖑 🏊 🚧 🤿 🎓 👘 TIC 🔜 📷				r
KNOWNS					UTP m/z: 482.9589-482.9637			
Compound Na	me Filter				,			
name *	m/z	rt î						
XANTHURE	204.030228	14.14	.3.6e+05					
XANTHOSIN		12.7		😣 💿 Peak Detection				+
XANTHOSINE	283.068420	7.74						T.
XANTHINE	151.026154	1.18		Feature Detection Selection Group Filtering Me	ethod Summary			1
VALINE	116.071701	1	3.2e+05					-
UTP	482.961273	14.81	.5.20+05	Current Peak Dection Method Summary				
URIDINE	243.062256	1.7		Save Method	Load Method			
URIC ACID	167.021057	6.17		SETTINGS		8		
URACIL	111.020004	1.18						0
UMP	323.028595	10	.2.8e+05	alignSamplesFlag=0				
UDP-N-ACE	606.074280	13.58		alignMaxItterations=10				
UDP-D-GLU	579.027039	14.68		alignPolynomialDegree=5				
UDP-D-GLU	565.047729	13.44		MASS SLICING				
UDP	402.994934	13.81	2.4e+05	rtStepSize=10				
TYROSINE	180.066620	2	1.40.00	massCutoffMerge=20				
TRYPTOPHAN	203.082596	7.5		limitGroupCount=1000				-
TREHALOSE	341.108948	1.18		minMz=0				
TREHALOSE		6.95		maxMz=1e+09		-		
TRANS TRA		16.74	.2e+05					
THYMINE	125.035652	2.6		Output Detected Peak to a Table				
	241.082993	5.4		Output Table	Create New Peak Table			
THREONINE		1						
THIAMINE-P		11	1.6e+05	Output Detected Peaks to CSV File				
THIAMINE P		9.56		Output Directory	16	Set Directory		
THIAMINE	264.105042	0.8		Output Directory		seconectory		
TAURODEO		16.23						
TAURINE	124.007393	0.82						
SUCCINYL-C		15.41	_1.2e+05	Done				
	117.019333	11.6			DOW EL AL	Peaks Cancel		
SN-GLYCER		7.3		10	500% Find	reaks Cancel		
SHIKIMATE		13.44						
	173.045547	10.92	8e+04		Use Default Parameters			
SERINE	104.035316	1			/ 82./~~ \$10.0			
SEDOHEPT		13.64						
SARCOSINE		1						
S-RIBOSYL-L		0.68						
S-METHYL-5'		11.6	_4e+04					
S-ADENOSY		7						
S-ADENOSY		1				A		
S-ADENOST	383 11/310	6						
S-MUENUST	363.114319				and the second se			

Peak Table

Align Peaks Match Options Polly		-		-			. ,	arge 1 : AreaTop		[m/z] 168.06654358		:) p
KNOWNS : 💕 🖕] 🙉 📌 😼 🕹 I	👾 🖛 🌩 🍺	TIC	- =								
Compound Name Filter						UTP m/z	: 482.9589-482.9637					
name * m/z rt	,3.6e+05											
(ANTHURE 204.030228 14.14												
ANTHOSIN 363.034729 12.7 ANTHOSINE 283.068420 7.74	_3.2e+05											
ANTHINE 151.026154 1.18								•				
ALINE 116.071701 1	.2.8e+05											
TP 482.961273 14.81 RIDINE 243.062256 1.7												
RIC ACID 167.021057 6.17	.2.4e+05											
RACIL 111.020004 1.18												
MP 323.028595 10 DP-N-ACE 606.074280 13.58	.2e+05											
DP-N-ACE 606.074280 13.58 DP-D-GLU 579.027039 14.68												
DP-D-GLU 565.047729 13.44	_1.6e+05											
DP 402.994934 13.81								6				
YROSINE 180.066620 2 RYPTOPHAN 203.082596 7.5	_1.2e+05											
REHALOSE 341.108948 1.18	1.20+05							2				
REHALOSE 421.075256 6.95												
RANS_TRA 381.123749 16.74 HYMINE 125.035652 2.6	_8e+04											
HYMINE 125.055652 2.6 HYMIDINE 241.082993 5.4							4					
HREONINE 118.050964 1	_4e+04											
HIAMINE-P 344.071350 11 HIAMINE P 424.037689 9.56							and the					
HAMINE P 474 037684 4 56	12981-0000		D () 13.81		14.31		14.81	10.51	The second s	10.81	16.31	
Table 1 l 🔗 🛠 🔋 🥻	🤌 뿣 🙀 🐞	1										-
	Observed m/z	Expected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		
N-acetyl-glutamate	188.0566	188.0564	13.11	0.11	16	16	12	9.9e+03	1230	0.85		
N-acetyl-glutamine	187.0724	187.0724	3.75	3.25	16	15	21	7.4e+04	491	0.84		
Indoleacrylic acid	186.0559	186.0560	7.80	7.14	16	16	17	1e+04	1145	0.85		
3-phosphoglycerate 3-phospho-serine	184.9857 184.0016	184.9857 184.0016	13.57 8.14	0.01	16 16	10 16	15 36	1.1e+04 3.7e+04	473 965	0.84 0.85		
tyrosine	180.0666	180.0666	4.60	2.60	16	3	44	9.6e+03	3	0.53		
tyrosine	180.0665	180.0666	3.26	1.26	16	4	29	1.5e+04	2	0.52		
tyrosine tyrosine	180.0665 180.0667	180.0666 180.0666	3.87 15.85	1.87 13.85	16 16	1 7	30 26	1.3e+04 8.9e+03	2 17	0.61 0.75		
	168.9907	168.9908	13.85	6.37	16	10	13	1.3e+04	777	0.83		
1 D-glyceraldehdye-3-phosphate 0 D-glyceraldehdye-3-phosphate 9 hydroxyphenylpyruvate	168.9907 168.9905 179.0351	168.9908	8.70 14.28	1.35 0.70	16 16	10 16	62 47	2.5e+04 7.4e+03	1175 23	0.84 0.82		

Groups information obtained after Peak Detection is stored and displayed in the form of a Peak Table with a row representing a group and its corresponding features in columns. Users can show/hide the peak table by clicking on



Peak Table Features

Following are the different features/columns in a peak table:

- *#*: is the serial number for a group
- *ID*: Group ID is assigned according to the search mode used during peak detection. In case of Automated search, groups are named by their m/z and retention time values separated by '@' sign. For example, ID for a group with 230.2 m/z and 1.89 RT will be given as '230.2@1.89'. In case of Database search, groups are annotated as a compound from the reference file. For example, 'malate'.
- *Observed m/z*: is the median m/z of the group.
- *Expected m/z*: is the m/z value provided in the reference file for the compound represented by the group. This field is populated only in case of Database search.
- *rt*: is the median retention time of the group.
- *rt delta*: is the difference between expected retention time from the reference file and the observed RT. This field is set to -1 in case of Automated Search.
- *#peaks*: is the number of peaks in the group.
- *#good*: is the number of good peaks in a group. A good peak is defined as one with its quality score above the defined threshold in Peak Detection dialog.
- *Max Width*: is the maximum peak width in a group. Peak width is defined as the number of scans over which a peak is spread.

- *Max AreaTop*: is the maximum peak AreaTop intensity in a group. AreaTop is one of the quantitation types used to represent peak intensity in El-MAVEN. Read more about the different quantitation types here.
- *Max S/N*: is the maximum peak signal/noise ratio in a group.
- Max Quality: is the maximum peak quality score in a group.
- Rank: is the group rank. The formula and parameters involved have been explained here.

Peak Table Menu Bar

n Align Peaks Match Options Polly	n 🗠 📲 💼 🗠 🛪	3. <u></u> G	t TIC									
	🖥 🙉 🛣 👼 🕹 Σ	# 🗕 🗕 🕨 🛛	IIC E			UTP m/s	: 482.9589-482.9637					
name ^ m/z rt	,3.6e+05											
XANTHURE 204.030228 14.14	1											
XANTHOSIN 363.034729 12.7 XANTHOSINE 283.068420 7.74	.3.2e+05											
ANTHOSINE 285.068420 7.74 ANTHINE 151.026154 1.18								•				
ALINE 116.071701 1	.2.8e+05											
JTP 482.961273 14.81												
JRIDINE 243.062256 1.7	2.4e+05											
JRIC ACID 167.021057 6.17 JRACIL 111.020004 1.18	.2.48+03											
JMP 323.028595 10												
JDP-N-ACE 606.074280 13.58	.2e+05											
JDP-D-GLU 579.027039 14.68												
JDP-D-GLU 565.047729 13.44	_1.6e+05											
JDP 402.994934 13.81 TYROSINE 180.066620 2												
ITROSINE 180.066620 2 IRYPTOPHAN 203.082596 7.5	_1.2e+05											
TREHALOSE 341.108948 1.18	1.20400							2				
TREHALOSE 421.075256 6.95												
TRANS_TRA 381.123749 16.74	_8e+04											
THYMINE 125.035652 2.6 THYMIDINE 241.082993 5.4												
THREONINE 118.050964 1	_4e+04						1					
FHIAMINE-P 344.071350 11							and the					
THIAMINE P 424 037689 9 56												
k Table 1 🔖 🥪 🕱 💡 ጅ 🤣		13.31			14.31		14.01	10.31		10.01	10.31	-
	Observed m/z	Expected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		
0 N-acetyl-glutamate	188.0566	188.0564	13.11	0.11	16	16	12	9.9e+03	1230	0.85		
9 N-acetyl-glutamine	187.0724	187.0724	3.75	3.25	16	15	21	7.4e+04	491	0.84		
8 Indoleacrylic acid	186.0559	186.0560	7.80	7.14	16	16	17	1e+04	1145	0.85		
7 3-phosphoglycerate	184.9857	184.9857	13.57	0.01	16	10	15 36	1.1e+04	473	0.84		
6 3-phospho-serine 5 tvrosine	184.0016 180.0666	184.0016 180.0666	8.14 4.60	0.06 2.60	16 16	16 3	36 44	3.7e+04 9.6e+03	965 3	0.85 0.53		
4 tyrosine	180.0665	180.0666	3.26	1.26	16	4	29	1.5e+04	2	0.52		
3 tyrosine	180.0665	180.0666	3.87	1.87	16	1	30	1.3e+04	2	0.61		
2 tyrosine	180.0667	180.0666	15.85	13.85	16	7	26	8.9e+03	17	0.75		
1 D-glyceraldehdye-3-phosphate	168.9907	168.9908	13.72	6.37	16	10	13	1.3e+04	777	0.83		
0 D-glyceraldehdye-3-phosphate 9 hydroxyphenylpyruvate	168.9905 179.0351	168.9908 179.0350	8.70 14.28	1.35 0.70	16 16	10 16	62 47	2.5e+04 7.4e+03	1175 23	0.84 0.82		

Multiple groups can be annotated as the same compound especially when retention time information is not used during Database search. The peak table provides options for filtering, comparing or exporting data from the table. Following are the different menu options available in the peak table:

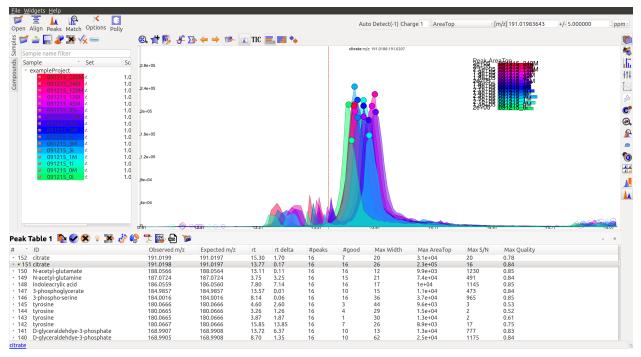


Switch between group and peak views: Switching to Peak view displays only Peak information. This includes group number, group ID, Expected m/z, Observed m/z, retention time and intensity of all peaks in the group with sample names as the respective column headers. Peak intensity cells are colored based on their relative values in a group. Highest intensity value has the lightest color and vice-versa.

	lgets <u>H</u> elp												
	Ξ 📐 🔐						A	uto Detect(-1) Charg	e 1 : AreaTop	: [m/z] 1	68.06654358	+/- 5.000000	: ppm :
	lign Peaks Match O		o 4 💼 10 –										rakena.
Samples Com	OWNS	: 📁 📊	🙉 🛣 👼 🕹 🏹	h 🔶 🤶 🔝	TIC	<u> </u>							r i
Leg Com	npound Name Filter						UTP m	/z: 482.9589-482.9637					N
		rt	,3.6e+05							Reak AreaTop	15 240M		1
	THURE 204.030228									6:38 1 82	15-728M		
& XAN	THOSIN 363.034729		.3.2e+05							2.86+65	15-45M		+++
5 XAN	THOSINE 283.068420								•		15-5M		N
- XAN	THINE 151.026154 INE 116.071701		2.8e+05						1	2:28:183	B 3M B IM		>
UTP			2.00100							7,38,105 18	B gM		
URI	DINE 243.062256	1.7								1.6e+05 0912	15_0I		C
	CACID 167.021057		.2.4e+05										۵,
URA													
	P-N-ACE 606.074280		.2e+05										R
	P-D-GLU 579.027039												
	P-D-GLU 565.047729		_1.6e+05										•
UDP	OSINE 402.994934 05INE 180.066620												0
	PTOPHAN 203.082596		_1.2e+05										
TRE	HALOSE 341.108948	1.18							6				
	HALOSE 421.075256		8e+04						5				
	NS_TRA 381.123749 MINE 125.035652		_00+04										
	MIDINE 241.082993												
	EONINE 118.050964		_4e+04										
	AMINE-P 344.071350 AMINE P 424.037689							and the					
(HIL	AVIINE P 474 037689	, 4 70 ,	291000	-13.31	© ()13:81		14.31	14.81	15.31	15.81	<u> </u>	16.31	
Peak T	able 1 🖍 🖌 🗶	💡 🐱 🤔 🍳	👂 梵 🙀 👘										- @
	ID		Observed m/z	Expected m/z	rt	091215 240M	091215 240i	091215_120M	091215 120i	091215 45M	091215 45i	091215 15M	09
	pyridoxine		168.0665	168.0666	3.22	3570.41	2426.26	4849.82	2334.67	4173.82	2559.66	4712.61	28!
· 175	phenylalanine		164.0716	164.0717	4.10	4.68463e+06	3.35967e+06	4.75213e+06	3.64234e+06	4.64550e+06	3.18198e+06	4.68504e+06	3.2
	Aminoadipic acid		160.0615	160.0615	4.82	1639.74 11474	688.027	821.929 17023	434.228	712.462	622.257	884.181	30
	Aminoadipic acid Aminoadipic acid		160.0613 160.0614	160.0615 160.0615	4.07 7.73	114/4 21976.6	6025.17 61080.8	21585.3	7892.2 64061.7	20120 20258.4	8345.76 57748.9	17736.2	76: 428
	2-Aminooctanoic acid		158.1187	158.1187	15.88	4442.94	4753.67	5514.79	4395.37	4650.03	5238.44	4511.24	45
· 170	methionine		148.0437	148.0438	3.47	5550.92	2373.85	2901.35	1462.92	3470.98	1529.74	2932.29	15
	methionine		148.0438	148.0438	13.13	2306.57	10058.5	2829.74	11319.6	1759.42	16674.2	1494.2	12:
	O-acetyl-L-serine O-acetyl-L-serine		146.0459 146.0458	146.0459 146.0459	5.80 7.73	16399.3 3983.08	5305.45 7716.51	17950.8 4358.98	5084.05 9836.51	20320.3 6773.02	10318.7 11924.3	19486.1 8598.33	62 98
	O-acetyl-L-serine		146.0458	146.0459	4.25	1.06525e+06	873441	1.33902e+06	973313	1.38217e+06	1.56378e+06	1.44338e+06	1.5
· 165	glutamate		146.0459	146.0459	5.80	16399.3	5305.45	17950.8	5084.05	20320.3	10318.7	19486.1	621
	alista ana ta		*** ****		-	2002.00	7747.54	10000	0000 54	1770 00	440343	0500.33	005°

~

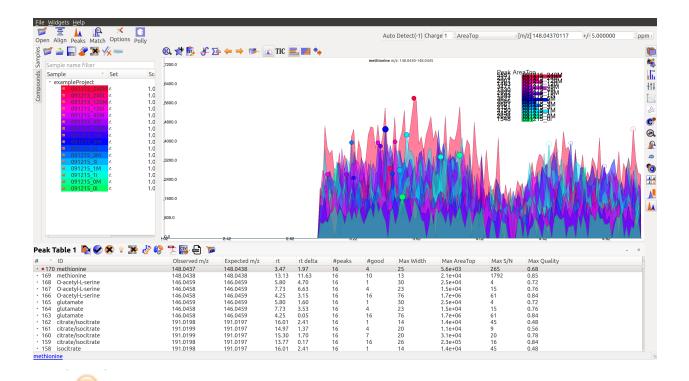
Mark Group as Good: Used to manually curate selected peaks as 'good'. User can also press 'G' on their keyboard for the same. Manual curation has been described here.





•

Mark Group as Bad: Used to manually reject peaks by marking them as 'bad'. User can also press 'B' on their keyboard for the same. Manual curation has been described here.



Train Neural Net: Used to retrain the neural net algorithm to recognize good/bad peaks. User manually curates 100 peaks to train the algorithm.

×

Delete Group: Deletes the selected group(s) from the peak table.

Show Scatter Plot: Opens the Scatter plot widget used to compare different cohorts via Scatter plot and Volcano plot.

The remaining are export options and will be detailed in the Export section.

Statistics

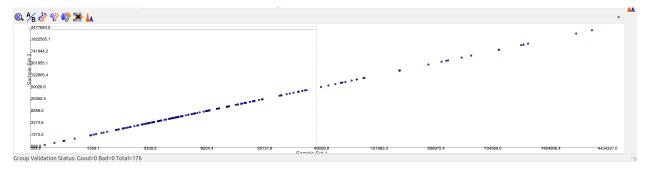
El-MAVEN comes equipped with a statistics module for comparing data across different cohorts. Users can set the sample cohorts either by editing the Set column in the Sample space, or upload a meta file with sample and cohort names as detailed above under the Sample Space Menu section.

The statistics module can be accessed through the Peak Table menu.

Align Peaks Match Options Polly													
NOWNS : 📁 🔚	🙉 😭 👼	∫ Σ≞	🔶 🔶 🃂 🔟 TIC		•								
mpound Name Filter	.9e+04		Compare Samples				UDP m/z: 40	2.9929-402.9969					
ime * m/z rt	8.1e+04		compare samples										
STATHIO 221.060150 1	3.10+04		Set 1					Set 2					
CLIC-AMP 328.045258 13.44 CLIC BIS(689.087585 14.2	.7.2e+04												
P 481.977234 14.9	6.3e+04											_	
EATININE 112.051636 EATINE 130.062195	b.30+04												
EATINE 130.062195 DENZYME A 766.107971 15.4	.5.4e+04											<u>}}</u>	
AP 322.044586 8.5													
TRULLINE 174.088409 0.9	_4.5e+04												
TRATE/IS 191.019730 13.6 TRATE 191.019730 13.6	.3.6e+04												
TRACONI 129.019333 13.29													
OLINE 103.100266 OLIC ACID 407.280304 16.69	,2.7e+04												
IOLESTER 465.304413 17.25	_1.8e+04												
LLOBIOSE 341.108948 3.86													
DP-ETHAN 445.053101 6.38 DP-CHOLINE 488.107880 6.33	_9e+03	0 sam	ples selected			0 samples	selected						
102 010000 10.55		Com	pare Sets Rese	t l	Done								6
	12.51	MinLO	G2 Fold Diff -100.00	- N	1in. Intensity	1.0		p. value	1.00000000	-14.	.01	14.91	
Bookmark Table 🍢 父 🛠 🤇) 🜁 🗋 🖑	Set Mir	sing Values 100.00		1in. Good Sai				n No Correction :				
# ^ ID	0	D:							P	Max S/N	Max Quality	r	
 176 pyridoxine 175 phenylalanine 		8.0665	168.0666 164.0717	3.22 4.10	1.42 0.30	16 16	4 16	31 83	5.9e+03 5e+06	4 126	0.56 0.84		
 174 Aminoadipic acid 		0.0615	160.0615	4.10	1.60	16	8	25	2.1e+04	925	0.82		
 173 Aminoadipic acid 		0.0613	160.0615	4.07	0.85	16	16	36	2e+04	1877	0.85		
 172 Aminoadipic acid 		0.0614	160.0615	7.73	4.51	16	16	16	6.4e+04	28	0.84		
 171 2-Aminooctanoic acid 170 methionine 		8.1187	158.1187 148.0438	15.88 3.47	-1.00 1.97	16	14 4	38 25	5.5e+03 5.6e+03	9 265	0.81 0.68		
 170 methionine 169 methionine 		8.0437	148.0438	3.47	1.97	16 16	4 10	13	2.1e+04	1792	0.68		
 168 O-acetyl-L-serine 		6.0459	146.0459	5.80	4.70	16	1	30	2.5e+04	4	0.72		
 167 O-acetyl-L-serine 		6.0458	146.0459	7.73	6.63	16	4	23	1.5e+04	15	0.76		
 166 O-acetyl-L-serine 		6.0458	146.0459	4.25	3.15	16	16	76	1.7e+06	61	0.84		
 165 glutamate 		6.0459	146.0459	5.80	1.60	16	1	30	2.5e+04	4	0.72		
 164 glutamate 		6.0458	146.0459	7.73	3.53	16	4	23	1.5e+04	15	0.76		
 163 glutamate 		6.0458	146.0459	4.25	0.05	16	16	76	1.7e+06	61	0.84		
 162 citrate/isocitrate 	19	1.0198	191.0197	16.01	2.41	16	1	14	1.4e+04	45	0.48		
 161 citrate/isocitrate 	19	1.0199	191.0197	14.97	1.37	16	4	20	1.1e+04	9	0.56		
							7				0.70		
 160 citrate/isocitrate 	19	1.0199	191.0197	15.30	1.70	16	7 16	20	3.1e+04	20	0.78		

- Set1/Set2: Select two cohorts to be compared
- *Min Log2 Fold Difference*: Fold difference is a measure of how much the intensity of a group changes from one cohort to another. User can set the minimum threshold for this value in log₂ format.
- *Min Intensity*: Groups with all peak intensities less than this value will be filtered out from the comparison process.
- *p value*: A t-test is performed to find if the intensity distributions of the two selected cohorts are significantly different from each other. This test returns a p-value indicating how significantly different a group behaves between the two cohorts. A lower p-value shows higher significance.
- Set Missing Values: User can set the default intensity value to be used in case the group is missing from a sample.
- *Min. Good Sample*: Groups should have a minimum number of good peaks (based on peak quality score) to be considered for comparison.
- *FDR Correction*: False discovery rate is the expected proportion of false positives in a test. There are a number of ways to correct for false positives. (Read More)
- Compare Sets: Click to get comparison results.

Compare Sets opens the scatter plot by default





Zoom Out: Zooms out of the plot.



Compare Samples: Opens the compare samples dialog again to adjust settings.

Scatter Plot: The axes represent the average peak intensity (Peak Height) for sample 1 and 2 respectively. Each bubble is a group. The bubble size represents fold change between the samples. The significance (or p-value) of the fold change is represented by the bubble color. Red and blue signify higher intensity in sample 1 and 2 respectively. Opaqueness of the bubble represents the significance (or inverse of p-value) of the fold change between cohorts.



Volcano Plot: The axes represent fold change and significance of fold change respectively. Red and blue bubbles represent positive and negative fold change respectively.



Delete: Deletes a data point from the graph.



Scatter Plot Table: A separate Peak Table is created with all filtered groups being used for statistical analysis. User can also export these in a CSV or JSON.

Export

Users can either save the state of the project or export only relevant data from the peak table. These are the different export options available in El-MAVEN:

<u>W</u> idgets <u>H</u> elp													
ad Samples Projects Peaks	Ctrl+O							Distant (d) cha			[1 5 000000	H.
ad Classification Model							Au	to Detect(-1) Cha	arge 1 : AreaTop	:	[m/z] 482.96166992	+/- 5.000000	i p
ad Compound List		o											
ve Project As	Ctrl+S 📊	🛛 🙉 🕁 👼 🕹 Σ	# 🗕 🍝 🖷	THC									
ve Settings							UTP m/s	: 482.9589-482.9637					
ad Settings													
otions		.3.6e+05								Reak Ar	ea 9915 120M		
ld S3 credentials										X:48183	213=338M		
port Bugs!										12.05			
it	Ctrl+Q	3.2e+05								8:58183	213-5EM		
XANTHINE 151.026154 VALINE 116.071701	1.18								•	3.38±85			
JTP 482.961273									T	5:48-04			
JRIDINE 243.062256	1.7	.2.8e+05								189A	AN SIZE AM		
URIC ACID 167.021057	6.17									1.00+00			
JRACIL 111.020004	1.18												
JMP 323.028595	10	_2.4e+05											
JDP-N-ACE 606.074280	13.58								1				
JDP-D-GLU 579.027039	14.68												
JDP-D-GLU 565.047729	13.44	_2e+05											
JDP 402.994934	13.81												
TYROSINE 180.066620 TRYPTOPHAN 203.082596	2 7.5												
TREHALOSE 341.108948	1.18	1.6e+05											
TREHALOSE 421.075256	6.95												
TRANS TRA 381.123749	16.74												
THYMINE 125.035652	2.6	1.2e+05											
THYMIDINE 241.082993	5.4												
THREONINE 118.050964	1												
THIAMINE-P 344.071350	11	Be+04											
HIAMINE P 424.037689	9.56												
HIAMINE 264.105042 AURODEO 498.289490	0.8							6					
AURINE 124.007393	0.82	_4e+04											
SUCCINYL-C 866.123962	15.41							A.A.					
SUCCINATE 117.019333	11.6							and the second					
· · · · · · · · · · · · · · · · · · ·		1281000		39 () (3.61		14.31		19.01	1000	and the second second	10.01	10.31	
c Table 1 🍢 🥪 寒	💡 🐱 🚀 (📎 梵 🚭 ե 🐞											- 0
* ID		Observed m/z	Expected m/z	rt	rt delta	#peaks	#good	Max Width	Мах АгеаТор	Max S/N	Max Quality		
5 glycerate		105.0193	105.0193	3.05	3.64	16	1	37	1.2e+04	18	0.74		
4 glycerate		105.0194	105.0193	5.54	1.15	16	14	29	9.5e+03	1079	0.84		
3 serine		104.0353	104.0353	3.56	2.56	16	2	26	6.1e+03	6	0.61		
2 4-aminobutyrate		102.0561	102.0561	5.78	1.78	16	0	30	9.8e+03	5	0.45		
 4-aminobutyrate dimethylglycine 		102.0560 102.0561	102.0561 102.0561	4.25 5.78	0.25	16 16	16 0	69 30	1.1e+06 9.8e+03	87 5	0.84 0.45		
10 aimecnyigiycine)9 dimethylglycine		102.0561	102.0561	4.25	-1.00	16	16	30 69	9.8e+03 1.1e+06	5 87	0.45		
s onliechyddydlie		102.0300	102.0301	4.20	-1.00	10	-		1.10+00	21	0.64		

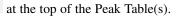
• Save Project as: This option is available in the File menu. It saves all peak tables and current settings in a .mzroll file. On loading the .mzroll file, all sample files are uploaded and the peak tables and EIC are available.

If the user wishes to save only certain Peak Tables, they can click on This will only store that specific peak table instead of all.

- Generate PDF Report: This option is available on their corresponding bar plots in a PDF file.
- Export Groups to SpreadSheet (.csv): This option is available on top of the Peak Table . You can choose to export the whole table or a subset of the data. There are 4 possible selections: export only selected groups, export all groups, export only good groups or export only bad groups. The data is stored in a comma separated file.







FIL	e <u>W</u> idgets <u>H</u> e														
0p	en Align Pea	aks Match Op	tions Polly						Au	to Detect(-1) Ch	arge 1 :AreaTop		[m/z] 482.96166992	+/- 5.000000	: ppm :
			: 💋 🔜	🙉 <table-cell-columns> 👼 🐇</table-cell-columns>	∑k 🔶 👄 🖡	🖈 TIC	= = 🐝								r
amp	• KNOWNS	lame Filter							UTP m/	z: 482.9589-482.9637					*
ds		* m/z	rt	_3.6e+05								Peak A	roaTop		<u> </u>
Compounds	XANTHURE	. 204.030228	14.14									169.4	1215 1218M		+†∔
du		363.034729 E 283.068420	12.7 7.74									43.2	181218 45M		
S	XANTHINE	151.026154	1.18	.3.2e+05								206.3	681215 1 51		\sim
	VALINE UTP	116.071701 482.961273	1 14.81								- Y	206.3 287:4	81215_3M		🚕
	URIDINE	243.062256	1.7	.2.8e+05								168:7	891213-0M		C?
	URIC ACID URACIL	167.021057 111.020004	6.17 1.18												
	UMP	323.028595	10	.2.4e+05											R
		. 606.074280 . 579.027039	13.58 14.68												
	UDP-D-GLU	. 565.047729	13.44	_2e+05											
	UDP TYROSINE	402.994934 180.066620	13.81 2												0
	TRYPTOPHAI	N 203.082596	7.5	_1.6e+05											
	TREHALOSE.	341.108948 421.075256	1.18 6.95												
	TRANS_TRA.	381.123749	16.74												
	THYMINE	125.035652 241.082993	2.6 5.4	_1.2e+05											
	THREONINE	118.050964	1												
		344.071350 424.037689	11 9.56	_8e+04							A				
	THIAMINE	264.105042	0.8												
	TAURODEO TAURINE	498.289490 124.007393	16.23 0.82	_4e+04											
	SUCCINYL-C.	866.123962	15.41							and a sta					
	SUCCINATE	117.019333	11.6	-		<u></u>							19.81		
Pe	ak Table 1	N	💡 💌 🤣 🖗	> 🏗 🔛 🕅											- 8
#	* ID			O Export Sel		rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		1
	115 glycerate			10 Export All	Groups 0193	3.05	3.64	16	1	37	1.2e+04	18	0.74		
	114 glycerate 113 serine	2		10 Export Goo 10 Export Bac		5.54 3.56	1.15 2.56	16 16	14 2	29 26	9.5e+03 6.1e+03	1079 6	0.84 0.61		3
+	112 4-aminot			102.0561	102.0561	5.78	1.78	16	0	30	9.8e+03	5	0.45		
	111 4-aminol 110 dimethyl			102.0560 102.0561	102.0561 102.0561	4.25 5.78	0.25 -1.00	16 16	16 0	69 30	1.1e+06 9.8e+03	87 5	0.84 0.45		
	109 dimethyl			102.0560	102.0561	4.25	-1.00	16	16	69	1.1e+06	87	0.84		
UT									^	~ '		-	A 4A		8
	• <i>F</i> _Y	nort Fl	Cs to 1	son [.] This	option is a	vail	able or	ton o	f the I	Peak Tab		It	exports all	EICs to	a Ison

EI-MAVEN Command Line Interface

file.

The peakdetector executable in El-MAVEN can load sample files, perform peak detection and save the resulting peaks in a CSV or mzroll file. It reads a configuration file with user-set parameters and paths to input and output files.

This executable is not part of installer versions of El-MAVEN. Please clone the repository and build El-MAVEN to use this feature.

Create a default configuration file

Users need an XML file to specify their configuration and specify directory paths. The XML template can be generated using a simple command and modified according to user requirements.

- Open the terminal
- Move to your local El-MAVEN folder:

\$ cd user@pc:~/ElMaven/ElMaven

- · Execute the following command
 - For Windows and Ubuntu:

\$./bin/peakdetector.exe --defaultxml

- For MacOS:

• A default XML will be created in your current folder

Peak detector parameters

CLI provides a limited number of parameters for user-modification. Default values are used for all other parameters used in El-MAVEN GUI.

The template XML contains three sets of parameters:

1. OptionalDialogArguments: Global parameters

- Users need to know the ionization mode and accepted values are 0 (neutral), -1 (negative) and +1 (positive): ionizationMode type="int" value="-1"
- Ionization charge. This is the number of positive or negative charge added to every ion. Accepted values are positive integers: charge type="int" value="1"
- Mass accuracy. Set the +/- ppm range for m/z values: compoundPPMWindow type="float" value="5"

2. PeaksDialogArguments: Peak detection parameters

• Peak Detection Algorithm. Accepted values: "0" for Compound Database Search, "1" (or higher) for Automated Feature Detection: processAllSlices type="int" value="0"

Following are the arguments used in Automated Peak Detection and their GUI equivalents. These can be found in the 'Feature Detection Selection' tab of Peaks Dialog:

- 'Limit m/z Range' minimum: minScanMz type="float" value="0"
- 'Limit m/z Range' maximum: maxScanMz type="float" value="1e+009"
- 'Limit Time Range' minimum: minScanRt type="float" value="0"
- 'Limit Time Range' maximum: maxScanRt type="float" value="1e+009"
- 'Limit Intensity Range' minimum: minScanIntensity type="float" value="50000"
- 'Limit Intensity Range' maximum: maxScanIntensity type="float" value="le+010"
- 'Mass Domain Resolution': ppmMerge type="float" value="20"

Following are the arguments used in Compound database search and their GUI equivalents. These can be found in the 'Feature Detection Selection' tab of Peaks Dialog:

- 'Select Database'. Enter full path to the database file: Db type="string" value="0"
- 'Match Retention Time (+/-)'. Value of '0' will set the flag as false. Positive integer value will set the allowed deviation in minutes: matchRtFlag type="int" value="0"
- 'Limit Number of Reported Groups per Compound': eicMaxGroups type="int" value="0"

Isotope detection. Following are the settings used to pull isotopes in case of database search and their GUI equivalents. These can be found in the 'Isotope Detection Options' in the Options Dialog:

• Choose labels present in your data. Enter a four digit value of '1's and '0's each representing a label: D2, N15, S34 and C13. For eg. 0000 signifies no labels, 0001 signifies your data has C13 label and 1111 signifies your data has all four labels: pullisotopes type="int" value="0"

EIC processing. Following are the parameters used for EIC processing and filtering and their GUI equivalents. These can be found under the 'Peak Detection' tab of Options Dialog:

- 'Maximum Retention Time Difference between Peaks': grouping _maxRtWindow type="float" value="0.5"
- 'EIC smoothing window': eicSmoothingWindow type="int" value="10"

Group Filtering. Following are the parameters used for filtering detected groups and their GUI equivalents. These can be found under the 'Group Filtering' tab of Peak Dialog:

- 'Minimum Peak Intensity': minGroupIntensity type="float" value="5000"
- Quantitation type of the intensity set in 'Minimum Peak Intensity'. Accepted values are- "0":AreaTop, "1":Area, "2":Height, "3":AreaNotCorrected: quantitationType type="int" value="0"
- 'At least x% peaks above minimum intensity' slider. Enter value in percentage. 0% will default to 1 peak: quantileIntensity type="float" value="0"
- 'Minimum Quality': minQuality type="float" value="0.5"
- 'At least x% peaks above minimum quality' slider: quantileQuality type="float" value="0"
- 'Minimum peak width'. Enter number of scans: minPeakWidth type="int" value="1"
- 'Minimum Signal/Baseline Ratio': minSignalBaseLineRatio type="float" value="2"
- 'Minimum Good Peak/Group'. Value should be less or equal to the number of samples: minGoodGroupCount type="int" value="1"
- 'Peak Classifier Model File'. The default model file can be found in the bin folder of El-MAVEN installation. Enter full path to the 'default.model' file: model type="string" value="0"
- Not used by CLI: rtStepSize type="float" value="20"

3. GeneralArguments: General parameters

- Alignment. Positive integer value will run an alignment algorithm: alignSamples type="int" value="0"
- Export EIC in JSON. Positive integer value will save EICs JSON in user-specifed output folder: saveEicJson type="int" value="0"
- Output directory. Enter full path to desired output folder: outputdir type="string" value="0"
- Save Project. Positive integer value will save the project as a .mzroll file in user-specified output folder. This file can be loaded in El-MAVEN GUI for further processing or visualization: savemzroll type="int" value="0"
- Sample Path. Enter full path to a sample file in each row: samples type="string" value="0"

Run

Once the parameters and directory paths have been set in the configuration file, run peak detection from the terminal using the following command from the El-MAVEN root directory:

• Windows and Ubuntu:

```
$ ./bin/peakdetector.exe --xml config.xml
```

• MacOS:

\$./bin/peakdetector.app/Contents/MacOS/peakdetector --xml config.xml

The resulting CSV file (and other files depending on the configuration) can be found in the specified output directory.

Help

To print the help commands, execute the following:

```
$ Peakdetector.exe -h
```

1.4.3 Tutorials

Unlabeled LC-MS Workflow

This is a tutorial for processing Unlabeled LC/MS data files through El-MAVEN.

Preprocessing

msConvert is a command-line/GUI tool that is used to convert between various mass spectroscopy data formats, developed and maintained by proteoWizard. Raw data files obtained from mass spectrometers need to be converted to certain acceptable formats before processing in El-MAVEN.

Input

msConvert supports the following formats:

- .mzXML
- .mzML
- .RAW ThermoFisher
- .RAW Walters
- .d Agilent
- .wiff ABSciex

The settings used for msConvert as a GUI tool are captured in the following screenshots:

🖳 MSConvertGUI (64-bit)		-	٥	×
List of Files File of file names File: Add Remove	About MSConvert			
	MS Level v Levels:			
	Add Remove			
	Filter Parameters			
Output Directory:				
Browse				
Options				
Output format: mzML V Extension:				
Binary encoding precision: 64-bit 32-bit				
Write index: Use zlib compression: TPP compatibility: Package in gzip:				
Use numpress linear compression:				
Use numpress short logged float compression:				
Use numpress short positive integer compression:				
Use these settings next time I start MSConvert			Star	rt

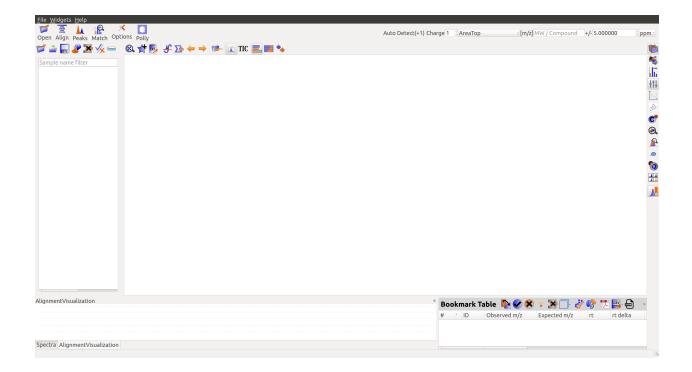
Note: zlib compression is enabled by default in msConvert. El-MAVEN in its current form does not support zlib compression. It is important to uncheck "Use zlib compression" box.

Output

msConvert can convert to an array of different formats but El-MAVEN primarily uses .mzXML and .mzML formats.

Launch EI-MAVEN

Once sample files are ready for processing, launch El-MAVEN.



Adjust Global Settings



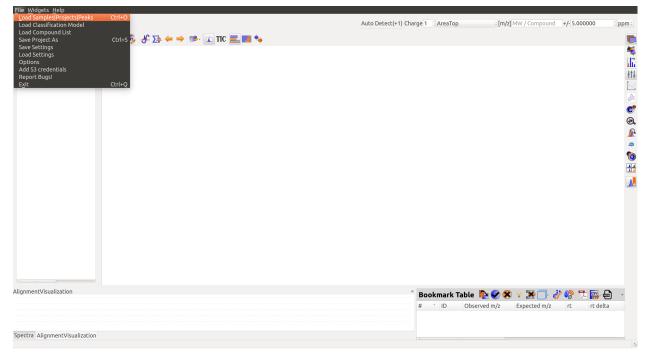
Global Settings can be changed from the *Options* dialog . There are 9 tabs in the dialog. Each of these tabs has parameters related to a different module in El-MAVEN. For example, a tab for Instrumentation details, a tab for File Import settings etc.

Open Align Peaks Match Options Poll	v		Auto Detect(+1) Charge 1	АгеаТор	: [m/z] MW / Compound +/	/- 5.000000
	🤊 👼 🚜 듲 🔶 🔝 TIC 🔜 🎫 🍫					l.
Sample name filter	Options Instrumentation File Import Peak Detection Peae Peak Detection and Grouping Settings EIC Smoothing Algorithm EIC Smoothing Window	ak Filtering Isotope Detectio	n EIC (XIC) [BETA] Peak Grouping Gi Baseline Calculation Drop top x% intensities from chror	· · · ·	er • i •	
	Max Retention Time difference between Peaks	s 0.25 i min Use Default Para	Baseline Smoothing	5 scans		
					_	
AlignmentVisualization			Boo	okmark Table - 📭 🤅	2 🙁 🔋 💌 🗍 🤣 🤅	> 📆 🐺 🖨
				1D Observed n	-	rt rt delta
Spectra AlignmentVisualization						

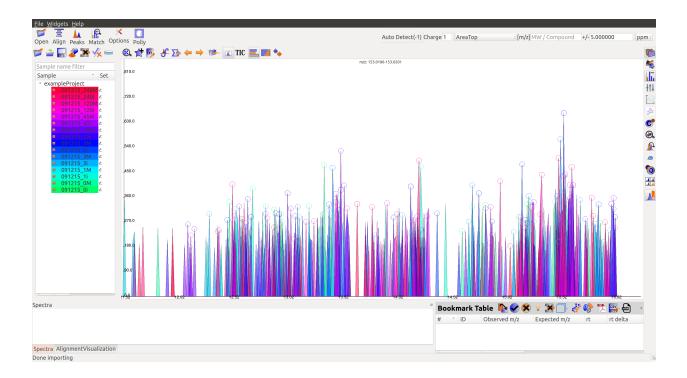
To know more about the functionality of different tabs and their settings, users can see the Widgets page. Please be sure to set the desired settings before processing an input file.

Load Samples

Users can go to File in the Menu and click on *Load Samples*|*Projects*|*Peaks* option. Then navigate to the folder containing the sample data, select all .mzXML or .mzML files and click Open. A loading bar displays the progress at the bottom.

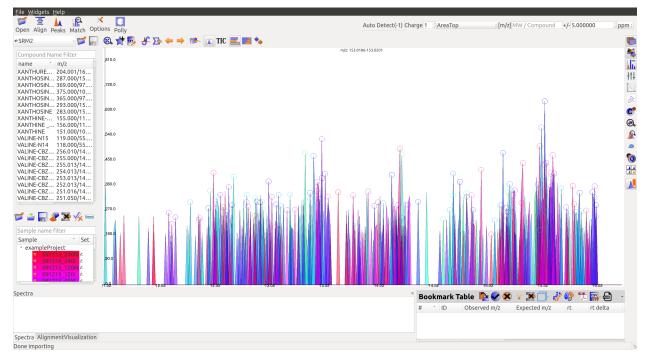


When the samples have loaded, users should see a sample panel on the left side. If it is not displayed automatically, click on the *Show Samples Widget* button on the toolbar. El-MAVEN automatically assigns a color to every sample. Users can select/deselect any sample by clicking the checkbox on the left of the sample name.



Load Compound Database

Users can click on *Compounds* option in the leftmost menu, navigate to the folder containing the standard database file, select the appropriate .csv file and click *Open*. Alternatively, users may use any of the default files loaded on start-up.



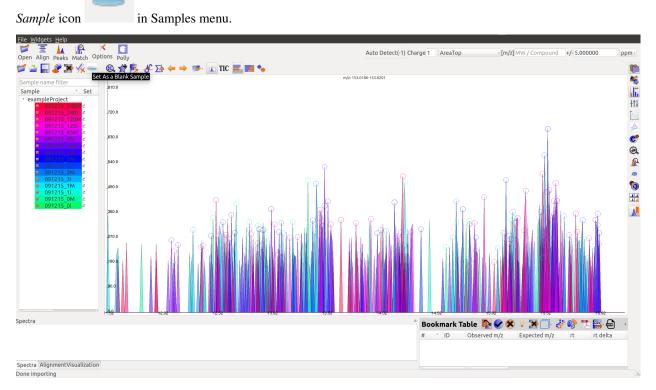
This is a sample Compound Database:

	A	В	С	D	E	F	G	н	1	J	ĸ	L	N
1	compound	formula	id	rt	category								
2	cholesteryl sulfat	C27H46O4S	HMDB00653	14.31									
3	Deoxycholic acid	C26H43NO5	C05464	0.82									
4	trans_trans-farne	C15H28O7P2	C00448	1									
5	Cholic acid	C24H40O5	HMDB00619	13									
6	Geranyl-PP	C10H20O7P2	C00341	16.46									
7	Taurodeoxycholic	C26H45NO6S	C05463	3.86									
8	lipoate	C8H14O2S2	C00725	8.5									
9	3-hydroxy-3-metl	C27H44N7O20P	C00356	1.89									
10	butyryl-CoA	C25H42N7O17P	C00630	7.45									
11	malonyl-CoA	C24H38N7O19P	C00083	6.7									
12	propionyl-CoA	C24H40N7O17P	C00100	7.7									
13	acetoacetyl-CoA	C25H40N7O18P	C00332	4.2									
14	acetyl-CoA	C23H38N7O17P	C00024	1.12									
15	dGTP	C10H16N5O13P	C00002	14.68	nucleic acids								
16	succinyl-CoA/me	C25H40N7O19P	C00091 C00683	0.7									
17	malonyl-CoA	C24H38N7O19P	C00083	11.6									
18	3-methylphenyla	C9H10O2	HMDB02222	7.35									
19	coenzyme A	C21H36N7O16P	C00010	14.14									
20	acetoacetyl-CoA	C25H40N7O18P	C00332	7									
21	Phenylpropiolic a	C9H6O2	HMDB00563	13.58									
22	3-hydroxybutyryl-	C25H42N7O18P	C01144 C03561	3.5									
23	guanosine 5'-dipl	C10H11N5O17P	4	15.64									
24	Indoleacoviic acid	C11H0NIO2	HMDR00734	7 5									

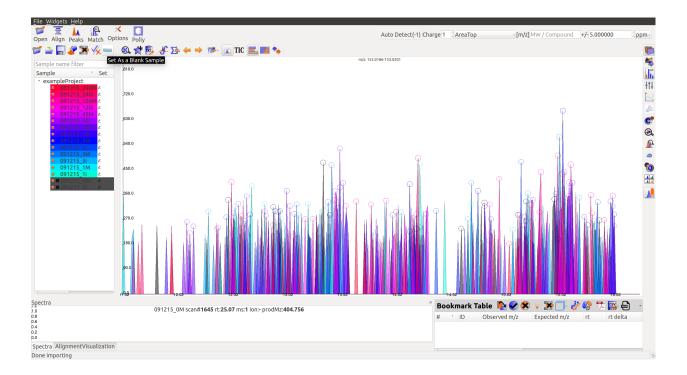
It lists all metabolite names, chemical formula, HMDB ID, and the class/category of compounds they belong to (if known).

Mark Blanks

Users can mark the blanks by selecting the blank samples from the list on screen, and clicking on the Set as a Blank

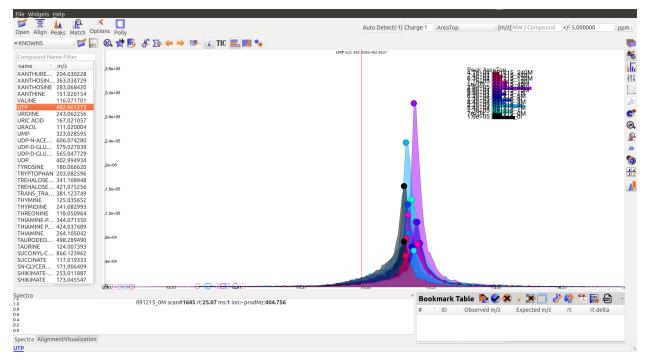


Multiple blanks can be marked together. The blanks will appear black as shown in the image below:

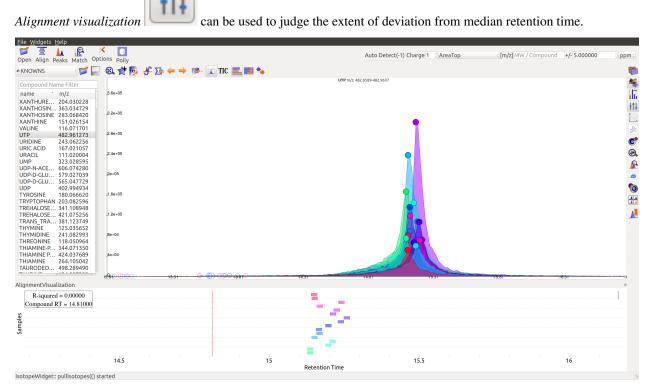


Alignment

Prolonged use of the LC column can lead to a drift in retention time across samples. Alignment shifts the peak retention time in every sample to correct for this drift and brings the peaks closer to median retention time of the group.



In the above image, EIC for a UTP group is displayed. If the samples were aligned, all peaks would lie at the same retention time. Since this is not the case, the samples need to be aligned.



In the above visualization, each box represents a peak from the selected group at its current retention time. Samples are said to be perfectly aligned when all peak boxes lie on the same vertical axis. The peaks are considerably scattered in the above image and therefore the samples should be aligned for better grouping of peaks.

Perform Alignment

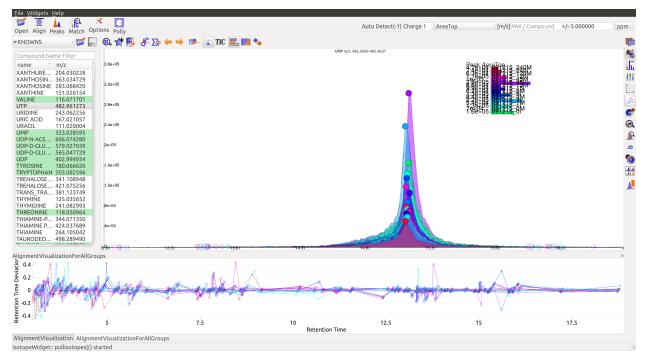
Alignment settings can be adjusted using the Align button rithm with default parameters.



. This example was aligned with Poly fit algo-

<u>F</u> ile <u>W</u> idgets <u>H</u> elp				
📁 🗄 📐 📭 🔅	< 💽	😣 💿 Alignment Options	e 1 : AreaTop : [m/z] MW / Compound	+/- 5.000000 . ppm :
Open Align Peaks Match Opt	ions Polly	Group Selection Criteria for use in Alignment		
≈KNOWNS : 📁 🔛	🙉 🙀 👼 🚜 🏊 👄 🔶 🔝 TIC			(m)
) = 10.0		Limit total number of groups in alignment to 1000 groups		
Compound Name Filter		Peak Grouping Window 10 scans		
name * m/z	3.6e+05		4.244 Atea 19215 240M	1
XANTHURE 204.030228		Alignment w.r.t Expected Rt	6:78183 881218=128M	+14
XANTHOSIN 363.034729 XANTHOSINE 283.068420	.3.2e+05	Peak Selection		111
XANTHINE 151.026154		Minumum Peak Intensity 5000.00	8:38±84 88 218=3M	\sim
VALINE 116.071701		Minimum Peak S/N Ratio 2		A
UTP 482.961273	_2.8e+05	Minimum Peak Width 3 scans	1380405 181218 AM	C°
URIDINE 243.062256 URIC ACID 167.021057			1.00+03 031213_01	
URACIL 111.020004	.2.4e+05	Peak Detection Algorithm Compound Database Search :		
UMP 323.028595		Select Database KNOWNS :		P
UDP-N-ACE 606.074280 UDP-D-GLU 579.027039	29+05			
UDP-D-GLU 565.047729		Alignment Algorithm		10
UDP 402.994934		Alignment Algorithm Poly Fit :		1
TYROSINE 180.066620	_1.6e+05	Maximum Number of Itterations 10		
TRYPTOPHAN 203.082596 TREHALOSE 341.108948		Polynomial Degree 5		
TREHALOSE 421.075256	_1.2e+05	Polynomial Degree 5		A.
TRANS_TRA 381.123749				
THYMINE 125.035652 THYMIDINE 241.082993	.Be+04			
THREONINE 118.050964	80+04			
THIAMINE-P 344.071350				
THIAMINE P 424.037689	_4e+04			
THIAMINE 264.105042 TAURODEO 498.289490				
TAORODEO 498.289490	1331			
AlignmentVisualization	12.612 (0.000) (0.00)		10.01	10.31
R-squared = 0.00000	1			
Compound RT = 14.81000				
Samples				
LEG				
		Status Undo Aligment		
		ondo Alginene		
	14.5	0% Align Cancel	5	16
AlignmentVisualization Alignmen	ntVisualizationForAllGroups			
IsotopeWidget:: pullisotopes(() st	arted			3

Post-alignment the peaks in the group should appear closer to the median retention time of the group.



File Widgets Help										1
Open Align Peaks Match Op						Auto Detect(-1) Charge 1	. AreaTop	: [m/z] MW / Compound	+/- 5.000000 . pr	pm :
] 🙉 🛠 👼 J 🗵	k ← → 🌮 [🗴 TIC 📃 🗾 🍫							Þ
Compound Name Filter					UTP m/z: 482.5	9589-482.9637				
name ^ m/z	3.6e+05						F	Peak AreaTopus		eniz .
XANTHURE 204.030228							2	:48187 88 1818_248M		ш
XANTHOSIN 363.034729								(16105 881213 43M		łtļ
XANTHOSINE 283.068420	_3.2e+05									To /
XANTHINE 151.026154						•	6			Ľ.
VALINE 116.071701	2.8e+05									≽
UTP 482.961273 URIDINE 243.062256	32.00100							RED AS BE S AM		C
URIC ACID 167.021057										U
URACIL 111.020004	2.4e+05									۵,
UMP 323.028595										R
UDP-N-ACE 606.074280										
UDP-D-GLU 579.027039	_2e+05									۲
UDP-D-GLU 565.047729 UDP 402.994934										1
TYROSINE 180.066620	1.6e+05									
TRYPTOPHAN 203.082596										.
TREHALOSE 341.108948										
TREHALOSE 421.075256	_1.2e+05									
TRANS_TRA 381.123749										
THYMINE 125.035652 THYMIDINE 241.082993	_Be+04									
THREONINE 118.050964	300+04									
THIAMINE-P 344.071350										
THIAMINE P 424.037689	_4e+04									
THIAMINE 264.105042										
TAURODEO 498.289490							All and a second			
······································	1281-000 0		2000 3.81	14.31		14.81	15.31	15.81 00 00 000	96.31	40
AlignmentVisualization										8
R-squared = -0.88222										
Compound RT = 14.81000					-					
es										
Samples				-						
S					4					
					<u></u>					
	14.5			15		15.5			16	
				Retentio	on Time					
AlignmentVisualization Alignme	entVisualizationForAllGr	oups								
IsotopeWidget:: pullisotopes(() s	started									8

Pre-alignment, the peaks were considerably scattered while the aligned peaks lie nearly on the same axis. Users can run alignment again with different parameters if required (or with a different algorithm). Further details on Alignment settings are available on the Widgets page.

Peak Grouping

Peak grouping is an integral part of the El-MAVEN workflow that categorizes all detected peaks into groups on the basis of certain user-controlled parameters. A group score is calculated for every peak during the process. The formula for this score takes into account the difference in retention time, intensities between peaks (smaller difference leads to a better score) and any existing overlap between them (higher extent of overlap leads to better score). All three parameters have certain weights attached to them that can be controlled by the users. The formula for the score is shown in the image. More details on it can be found on the Widgets page.

(😕 🗊 Options		
	Instrumentation File Import Peak Detection Peak Grouping Score score = 1.0/((distX*A)+0.01)/ ((distY*B)+0.01)*(C*overlap) where distX = Rt difference of peaks	Peak Filtering Isotope Detection EIC (XIC) [BETA] Peak Grouping Group Rank Libraries Exter • Consider distX Weight Other 1 Overlap Weight Overlap Weight 1 	
		Use Default Parameters	

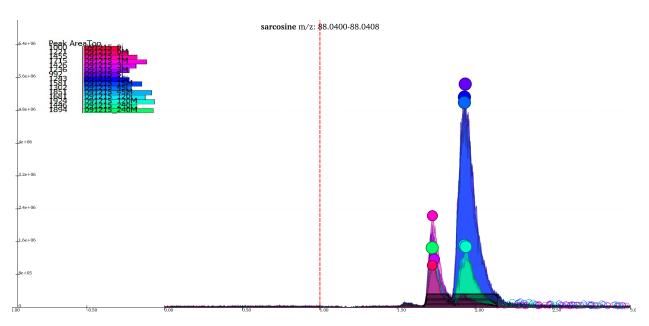
		sarcosine m/z: 88.04	400-88.0408		
_6.4e+06	Peak AreaTop 2.4e+05 AMPLE #SRK2BG7 2.2 3.9e+05 SMPLE #SRK2BG7 2.3 2.7e+05 SAMPLE #SRK2BG7 2.4 5.1e+06 SAMPLE #SRK2BG7 2.5 4.9e+06 SAMPLE #SRK2BG7 2.6 4.4e+06 SAMPLE #SRK2BG7 2.8 1.4e+06 SAMPLE #SRK2BG7 2.9 1.2e+06 SAMPLE #SRK2BG7 2.9 1.2e+06 SAMPLE				
5.6e+06	4.9e+06 SAMPL #SRK2BG7 2.6 4.4e+06 SAMPLE #SRK2BG7 2.7 1.4e+06 SAMPLE #SRK2BG7 2.8 1.4e+06 SAMPLE #SRK2BG7 2.9 1.2e+06 SAMPLE#SRK2BG7 2.10			2	
4.8e+06					
4e+06					
_3.2e+06					
_2.4e+06					
_1.6e+06				8	
8e+05					
0	l _{.0.50} 0.00	0.50	1.50	2.00	

The above image shows two groups in the EIC window. The highlighted (solid circles) peaks belong to group A, the peaks to its left with empty circles belong to another group B. The short peaks in group A that are close to the baseline and peaks in group B come from the same samples. Additionally, the high intensity peaks of group A have a similar peak shape to group B peaks. These peaks might have been wrongly classified into separate groups because of the difference in retention time range of the two sets of peaks. The weights attached to difference in retention time and intensities, and extent of overlap can be adjusted for better grouping.



Grouping parameters can be changed from the Options dialog

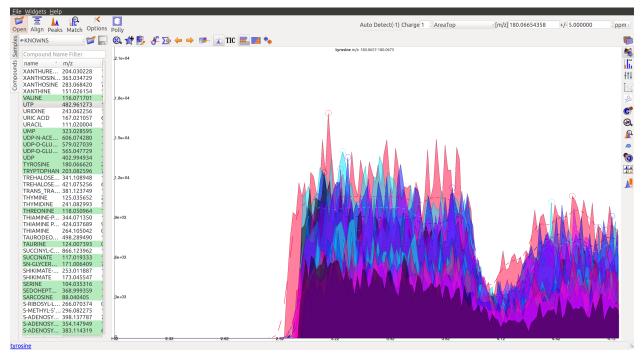
Options Instrumentation File Import Peak Detection Peak Grouping Score score = 1.0/((distX*A)+0.01)/(((distY*B)+0.01)*(C*overlap)) where distX = Rt difference of peaks	Peak Filtering Isotope Detection E Consider distX Weight distY Weight Overlap Weight	EIC (XIC) [BETA] Peak Group 0 0 1	ing Group Rank Libraries Exter (
	Use Default Paramet	ters	



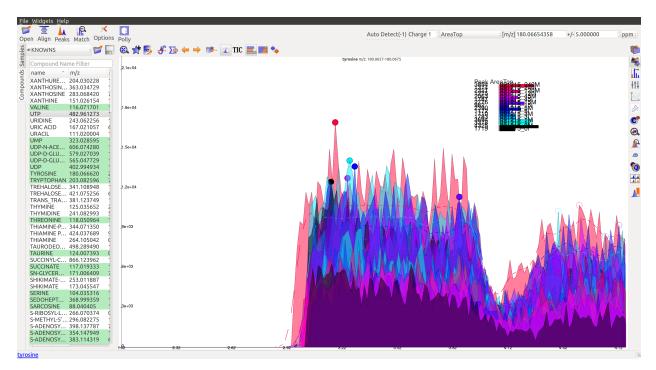
Giving less priority to difference in retention time and intensities results in the two groups being merged into a single group while the peaks that lay close to the baseline are no longer classified as valid peaks.

Baseline

When measuring a number of peaks, it is often more effective to subtract an estimated baseline from the data. This baseline should be set where ideally no peaks occur. Although sometimes the program sets a particular baseline such that one or more peaks occur below that baseline value. In the following image, the dashed line represents each baseline:



The corresponding peaks are indicated with solid circles:



The baseline correction can be done in the *Peak Detection* tab by clicking on *Options* button:

1	🔊 🗊 Options		_							
4	Instrumentation File Import Peak Detection Pe	ak Filtering Isotope De	etectior	EIC (XIC) [BETA] Peak Grouping Group Rank	Libraries Exter					
	Peak Detection and Grouping Settings		E	Baseline Calculation						
	EIC Smoothing Algorithm	Savitzky-Golay :		Drop top x% intensities from chromatogram	2%					
1	EIC Smoothing Window	10 1 s	scans							
	Max Retention Time difference between Peak	s 0.25 . n	nin	Baseline Smoothing	5 scans					
	Use Default Parameters									

Further details on settings can be accessed here.

The m/z option scans the groups to find any specific m/z value and plot its corresponding EIC. The +/- option to its right is to specify the expected mass resolution error in parts per million (ppm).

<u>F</u> ile	<u>W</u> idgets <u>H</u> e	lp								
1									a set of the set of the	
On		ks Match Options					Auto Detect(-1) Charge	1 AreaTop	: [m/z] MW / Compound +/	- 5.000000 ppm :
									-	
ole	KNOWNS	: 📁 📩) 🕴 🏹 🔊	🚽 🖓 🗢 🔿 🍽 🔟	TIC 📃 🗾 🍫					1
						UTP m/z	482.9589-482.9637			
	Compound N									
Compounds		m/z								*
INO	XANTHURE		_3.6e+05					R	BAK ANGATABIS 240M	+†+
đ	XANTHOSIN.							6	38183 881 213-728M	TI+
ē	XANTHOSINE XANTHINE	151.026154						j.		\sim
	VALINE	116.071701 1						9	98484 884213-13	
	UTP	482.961273	_3.2e+05					9. 8		2
	URIDINE	243.062256						Ş		©* @
	URIC ACID	167.021057 🤅					T T	7	6e465 891213 8M	õ
	URACIL	111.020004								
	UMP	323.028595	_2.8e+05							<u></u>
	UDP-N-ACE UDP-D-GLU									
	UDP-D-GLU									
	UDP	402,994934	2.4e+05							0
	TYROSINE	180.066620 2	.2.46403				9			
	TRYPTOPHAN									
	TREHALOSE						M			
	TREHALOSE		2e+05							
	TRANS_TRA THYMINE	125.035652 2								
	THYMIDINE									
	THREONINE									
	THIAMINE-P	. 344.071350	1.6e+05							
	THIAMINE P									
	THIAMINE	264.105042 (
	TAURODEO TAURINE	124.007393								
	SUCCINYL-C		1.2e+05							
	SUCCINATE									
	SN-GLYCER									
	SHIKIMATE									
		173.045547	.8e+04							
	SERINE	104.035316								
	SEDOHEPT SARCOSINE									
	S-RIBOSYL-L.									
	S-METHYL-5'.		_4e+04							
	S-ADENOSY	. 398.137787					ALD /			
	S-ADENOSY						A BARA			
	S-ADENOSY	. 383.114319 🤅			~~					
			12/81 (000)	13:31	0.013.81	14.31	114.01	10.31	15.81 (2016.3)	

Mass Spectra



Mass Spectra Widget displays each peak, its mass, and intensity for a scan. As the widget shows all detected masses in a scan, the ppm window for the EIC and consequently grouping can be adjusted accordingly. This feature is especially useful for MS/MS data and isotopic detection.

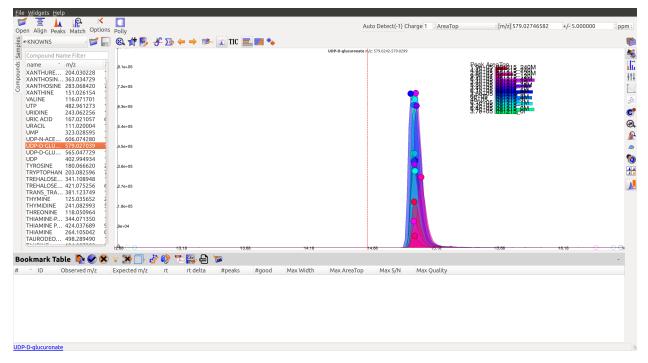
<u>File</u> <u>W</u> idge														
12 🗐	🛕 📭	×							Auto Detect(-1) Ch	arge 1 : AreaTop		[m/z] MW / Compound	+/- 5.000000	: ppm :
	n Peaks Match													
Samples Combo	NS : 1	📁 🖳 I	🙉 📌 👼	, 🕹 🏹 🔶 🔶	🣂 主 TIC 💻	, 🥅 🍫								6
un a	the with							UTP m/z: 4	82.9589-482.9637					*
Compo														
name XANTH XANTH XANTH XANTH	^ m/z		_3.6e+05								4784A4	1315 340M		dü
S XANTH	URE 204.03022										6:78:83	881218-128M		+++
E XANIH	OSIN 363.03472 OSINE 283.06842		3.2e+05								2 88185	2 2 2 2 2 M		117
XANTH											8:28:84	1212-12-12M		\sim
VALINE										T	2:28784	881213 3M		۵
UTP	482.9612		_2.8e+05							A	738745			
URIDIN											1.66+05	091215_01		C
URIC AG			2.4e+05											C * @
UMP	323.02859		1											~
	ACE 606.07428													R
	GLU 579.02703		.2e+05											
	GLU 565.0477													
UDP TYROSI	402.99493 NE 180.06663		1.6e+05							0				0
	OPHAN 203.08259		1.00105											<u></u>
	LOSE 341.10894													
TREHAI	LOSE 421.0752	56 (_1.2e+05											
	TRA 381.12374									O				
THYMIN			0.04											
	DINE 241.08299 NINE 118.05090		_8e+04							C 1				
	NE-P 344.0713													
	NE P 424.03768		4e+04											
THIAMI														
TAURO	DEO 498.28949	90 1												
			129310 (1000)	43.31	- (COD) - G aar		14:31		19.01	10.31		10.81	963)	
Spectra														
3.6e+05					091	215_45i scan# 998	rt: 15.19 m	s:1 ion:- prod	Mz:482.961					
3.2e+05					482.961395 ^{z=1}									
2.8e+05					402.501355									
2.4e+05														
2e+05														
1.6e+05														
1.2e+05														
8e+04														
4e+04							403	964752						
Bo.96	481	66	481.977600	482.36	483.06		83.76	904752	484.46	484.965759 16		485.86485.972137	486.56	
	et:: pullisotopes(100100					100110				3

Peak Curation

There are multiple ways to curate peaks in El-MAVEN, though following are the two broad workflows:

1. Manual Peak Curation using Compound DB widget

To use manual curation using the compound DB widget, users have to iterate over all the compounds in the compound DB on the extreme left of the window, as highlighted in the image below.



Once on a compound, El-MAVEN shows the highest ranked group for that m/z. Users can now choose a group or reject it.

First, users need to double click on the peak group of their choice. This will get the retention time line to the median of the group and also add the metabolite to the bookmarks table (as shown in the image below). Users can read more about the bookmarks table here.

<u>File W</u> idgets <u>H</u> e														
📁 \Xi 💧		•							Auto Detect(-	1) Charge 1	АгеаТор	: [m/z] 579.02746582	+/- 5.000000	: ppm :
	ks Match Options													
Sample KNOWNS Compound N	: 📁 🗖	🖏 🛒	👼 子 泽 🔶		TIC	E 14 %								()
Compound N	ame Filter							UDP-D-glu	curonate m/z: 579.0242-579.	.0299				*
	m/z	8.1e+05									Pe	AKAKEATABIS 240M		esta
S XANTHURE	. 204.030228										4:	18113 181313 738M		+++
E XANTHOSIN.	363.034729 283.068420 7	7.2e+05									6010			TI+
XANTHINE	151.026154	1.20+05								•				\sim
VALINE	116.071701										26			۵
UTP URIDINE	482.961273 243.062256	_6.3e+05									5	98783 891213 8M		C
URIC ACID	167.021057 (_		ě
URACIL	111.020004 323.028595	5.4e+05												>> © • •
	. 606.074280													<u></u>
	. 579.027039	4.5e+05												۵
UDP-D-GLU	. 565.047729 402.994934													1
TYROSINE	180.066620 2	3.6e+05												
	N 203.082596 7													
TREHALOSE.	421.075256 🤅	2.7e+05												
	381.123749													
THYMINE	125.035652 2 241.082993 5	1.8e+05												
THREONINE	118.050964	1.00.00												
	. 344.071350	9e+04								O				
THIAMINE	264.105042	196+04												
TAURODEO	. 498.289490													
		12.68	13	.18	,	13.68	4,	1.18	14.68	01		15.68	16.18	- 0 0 1
Bookmark Ta	ble 💽 🖌	💡 🔀 🗍]- 🤣 🤣 🎵	📴 - 🔤	1									
# ^ ID	Observed		Expected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality			
1 UDP-D-g	lucuronate 579.0275		579.0270	15.05	0.37	16	16	14	6.4e+05	32	0.84			
Group Validation S	Status: Good=0 Bad=0) Total=1												8

When the users select the first group, they would be asked if they would like to auto-save the state of the application. This feature allows the users to go back to their curated peaks if they so wish in future.

1		P. ×								Auto Detect(-	1) Charge 1	АгеаТор	: [m/z] 168.06654358	+/- 5.000000	: ppm :
		ks Match Options		👼 🕹 🏹 🔶	-	- 🗼 TIC 📕									r
s Samp	KNOWNS	ame Filter							UDP-D-glu	curonate m/z: 579.0242-579	.0299				×.
UN X	ANTHURE	m/z 204.030228 . 363.034729	.8.1e+05												111. ††‡
· · · ·		283.068420 151.026154 116.071701	_7.2e+05								2				
L	JTP JRIDINE JRIC ACID	482.961273 243.062256 167.021057	_6.3e+05												
L L	JRACIL JMP	111.020004 323.028595	.5.4e+05												<u>ک</u>
L	JDP-D-GLU JDP-D-GLU		_4.5e+05												
Т	JDP 'YROSINE 'RYPTOPHAN	402.994934 180.066620 203.082596	.3.6e+05					Autosa Ø Autosa Ø Do yo	ve ou want to enabl		2				%
Т	REHALOSE	. 341.108948 . 421.075256 . 381.123749	.2.7e+05					autos	ave?	Yes	1				
T T	HYMINE HYMIDINE	125.035652 2 241.082993 5 118.050964	_1.8e+05					_							
Т	HIAMINE P	. 344.071350 . 424.037689 264.105042	_9e+04								2				
Т		498.289490	12988	13.1	8	13.6	18	74	.18	14.68	<u> </u>	.18	15.68	-18.18	
Boo	kmark Tal	ble 陷 冬 😣	💡 ጅ 🗍] 🤣 🧐 🏌	🛒 🖶	1									-
#	1 ID	Observed ucuronate 579.0275		Expected m/z 579.0270	rt 15.04		#peaks 16	#good 16	Max Width 14	Max AreaTop 6.4e+05	Max S/N 32	Max Quality 0.84			
1	טאייסטע	uculonale 579.0275		519.0270	15.04	0.36	10	10	14	0.40+05	32	0.04			
Group	Validation SI	tatus: Good=0 Bad=0) Total=1												8

Qualifying peaks as good or bad is explained in the next few sections.

2. Automated Peak Curation

El-MAVEN can automatically select high intensity and high quality groups. This workflow is called automatic peak curation. To enable this workflow users have to click on the peak detection widget present in the top left of the window.

Upon clicking the peak detection widget

the following dialog box will open.

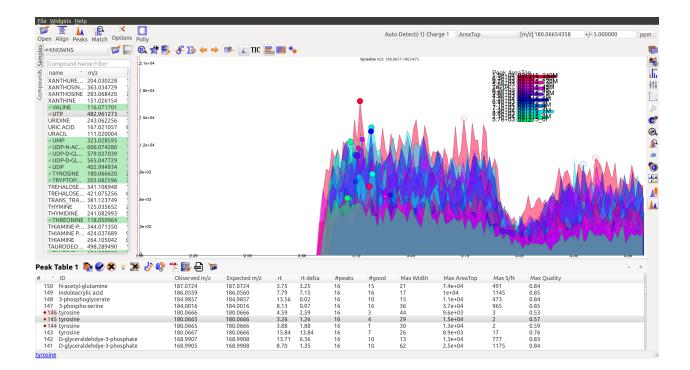
Instrumentation File Import Peak Detection Peak Filtering	Isotope Detectio	n EIC (XIC) [BETA] Peak Grouping Group Rank	Libraries Exter 🕐								
Peak Detection and Grouping Settings		Baseline Calculation									
EIC Smoothing Algorithm Savitzky-G	Golay :	Drop top x% intensities from chromatogram	80%								
EIC Smoothing Window 10	: scans										
Max Retention Time difference between Peaks 0.25	: min	Baseline Smoothing	5 scans								
Use Default Parameters											

Users can read more about the peak detection widget here.

Upon selecting the default parameters, users can click on *Find Peaks* to select the most important peaks. Once the peak detection is completed, a peak table shows up at the bottom of the window.

a Align Peaks Match Options	Polly						Au	to Detect(-1) Cha	arge 1 : AreaTop	: [m/z] 565.04821777	+/- 5.000000	
KNOWNS : 📂 🔙		ן 🛃 🖓 🔶	🣂 👔 TIC 🚪	2 📰 🍕	•								
ompound Name Filter	9e+05					UDI	P-D-glucose m/z:	565.0449-565.0505					
ame m/z	.9e+05									Peak Area	19815 040M		
ANTHURE 204.030228										2:38784 8	1215-720M		
ANTHOSIN 363.034729	_8.1e+05									128183	1213 45M		
ANTHOSINE 283.068420 7 ANTHINE 151.026154										3 48 1 85	215-5M		
VALINE 116.071701	.7.2e+05						7						
UTP 482.961273										2.88183	S S GM		
RIDINE 243.062256 1 RIC ACID 167.021057 (6.3e+05									/@+05** 🚺	91215_0		
RACIL 111.020004 1													
UMP 323.028595 1	5.4e+05												
UDP-N-AC 606.074280													
UDP-D-GL 579.027039 1 UDP-D-GL 565.047729	4.5e+05												
UDP 402.994934							81						
TYROSINE 180.066620 2													
TRYPTOP 203.082596 7 REHALOSE 341.108948	_3.6e+05												
REHALOSE 341.108948													
RANS_TRA 381.123749	_2.7e+05												
HYMINE 125.035652 2													
HYMIDINE 241.082993 5 THREONINE 118.050964	1.8e+05												
HIAMINE-P 344.071350													
HIAMINE P 424.037689 5	9e+04												
HIAMINE 264.105042 (AURODEO 498.289490													
AURODEO 498.289490													
	12:24	12.54		12.84		13.14		\$3.44	13.74		14.04	14.34	
Table 1 🖌 🖌 🏌	🛎 🤣 😹	梵 🚭 🐺											-
ID		Observed m/z	Expected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		
UDP-D-glucose		565.0482	565.0477	13.33	0.11	16	16	15	7e+05	32	0.84		
GTP		521.9838	521.9834	15.29	0.35	16	16	93	3.9e+04	949	0.84		
ATP UTP		505.9888 482.9617	505.9885 482.9613	15.31 15.18	0.37 0.37	16 16	16 16	61 91	1.1e+06 2.8e+05	11 109	0.82		
CTP		481.9777	481.9772	15.03	0.13	16	16	90	4.4e+04	2139	0.84		
dTTP		480.9825	480.9820	15.35	0.41	15	8	18	1.1e+04	562	0.84		
NAD+		662.1021	662.1019	9.14	0.52	16	16	32	1.2e+05	32	0.84		
glucose-6-phosphate		259.0222 259.0222	259.0225 259.0225	7.87 8.37	1.02 1.52	16 16	11 11	27 47	1.1e+04 7.5e+03	971 718	0.81 0.83		
glucose-6-phosphate													

Users can now iterate over these peaks by marking them as good or bad by clicking on the good or bad buttons present in the peak table as shown below.

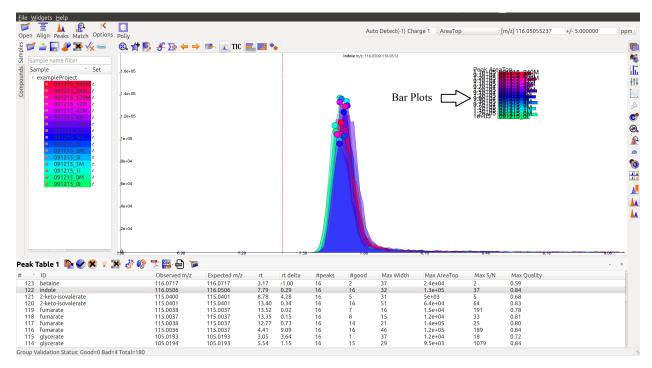


Guidelines for Peak Picking

• A peak's width and shape are two very crucial things to look at while classifying a peak as good or bad. A peak's shape should have a Gaussian distribution and width should not be spread across a wide range of retention time.

1	idgets <u>H</u> elp							,	Auto Detect(-1) Cha	Irge 1 : AreaTop	4	[m/z] 116.05055237	+/- 5.000000	: ppm :
Open Samples Sar	Align Peaks Match Options		, チ∑+ ← →	🃂 😰 TIC 📕		•								
ame								indole m/z: 1	16.0500-116.0512					
vi Sar	nple name filter											-		
	mple * Set	1.6e+05									Lage at the second			
noc	exampleProject ■ 091215 240M <										5:78=18			+1
Ĩ.											9:28 1 8	81215 15M		
S	✓ 091215 120M	1.4e+05					<u>(</u>				138182			C
	Ø 091215_120i							6			9:38±18			2
	<mark>◎ 091215_45M</mark> <	1.2e+05									1 28485	81213 AM		
	Ø91215_45i Ø91215_15M	1.20103									16+03	01210_01		C
	■ 091215_15M ≥						2	9						
	■ 091215 SM <	1e+05												
	🖉 091215_5i 🖉													
	■ 091215_3M <								Gaussian I	Distibution				4
	■ 091215_3i ■ 091215_1M	8e+04							Gaussian I	Jistioution				
	■ 091215_1M ≥													6
	■ 091215 0M <													-
	🖉 091215_0i 🖉	.6e+04					<u>M</u>							
							11							
														4
		_4e+04												
		_2e+04												
		3.00	0.90	V.20		4 9.06		V.00		0.10	0.40	0.70	9.0	
Peak '	Fable 1 🖉 🖉 🌹	🗶 🚀 🦉	📾 🔛				' W	Vidth						- 8
# ^	ID		Observed m/z	Expected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		
	betaine		116.0717	116.0717	3.17	-1.00	16	2	37	2.4e+04	2	0.59		
122	indole		116.0506	116.0506	7.79	0.29	16	16	32	1.3e+05	37	0.84		
	2-keto-isovalerate		115.0400	115.0401	8.78	4.28	16	5	31	5e+03	5	0.68		0
	2-keto-isovalerate		115.0401	115.0401	13.40	0.34	16	16	51	6.4e+04	64	0.83		
	fumarate fumarate		115.0038 115.0037	115.0037 115.0037	13.52 13.35	0.02 0.15	16 16	7 8	16 15	1.5e+04 1.2e+04	191 33	0.78 0.81		
	fumarate		115.0037	115.0037	13.35	0.15	16	8 14	21	1.2e+04 1.4e+05	25	0.81		
	fumarate		115.0036	115.0037	4.41	9.09	16	16	46	1.2e+05	189	0.80		
	glycerate		105.0193	105.0193	3.05	3.64	16	1	37	1.2e+04	18	0.72		
	glycerate		105.0194	105.0193	5.54	1.15	16	15	29	9.5e+03	1079	0.84		
C	alidation Status: Cood-0 Bad	A Tabal 400												

• Peak Intensities for a group are plotted as bar plots for all the samples. These bar plots have heights relative to the other samples. Thus, for a good peak the intensities should be high.



- Intensity Barplot heights should be higher for all the samples than Blank samples. We use intensities of Blank samples to set our group baseline. Blank intensities are used to calibrate intensity values across zero concentration.
- A good peak should have standards with varying intensity in a particular fashion (increasing or decreasing).
- Quality Control (QC) samples give us information about the quality of the data, i.e., it assesses reproducibility and software performance. Samples whose intensities and concentrations are already known are used as QCs to determine if the instrument is working as expected. Values (and scales) can be calibrated using QCs.
- If peak groups of a particular metabolite are separated apart (not aligned well) then we should use stringent alignment parameters to overcome this problem.
- For a particular metabolite, let's say if it has n number of groups, then the group which is much closer to the above guidelines should be selected as a good peak. Multiple groups can also be selected in case of ambiguity (if retention time information is not provided).

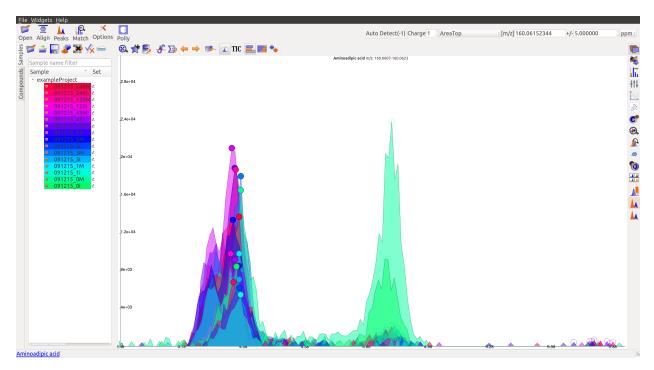
A good peak would look similar to the following peaks:

<u>File Widgets H</u> elp				
Open Align Peaks Match Option:	S Polly	Auto Detect(-1) Charge 1 AreaTop	: [m/z] 148.04383850	+/- 5.000000 ppm :
월 📁 🧀 🔛 🏖 🗏 🌾 😑	28+04			Ö
Sample name filter		methionine m/z: 148.0430-148.0445		×6
පු Sample * Set				
exampleProject 091215_240M	2.8e+04		2334 AS 379315_340M	+1+
© 091215_240i			2 301-03 101215-120M	W.
◎ 091215_120i			1:50-03 881212 12M 3:30-03 881212 5M	
<mark>∞ 091215_45M</mark>	2.4e+04			C ⁹
Ø 091215_15M ≤ Ø 091215_15i ≤		T T	291213_8M	0
<mark>∞ 091215_5M</mark>		M		₩ 11 11 2 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
■ 091215_3M <	20+04			۵
■ 091215_3i < ■ 091215_1M <				
Ø 091215_1i Ø 091215_0M				1
<mark>∞ 091215_0i </mark> «	1.6e+04			
	1.00+04			
				A
				-
	1.2e+04			
	8e+03			
	_4e+03			
· · · · · · · · · · · · · · · · · · ·			13.83	
IsotopeWidget:: pullisotopes(() starte	ed			8

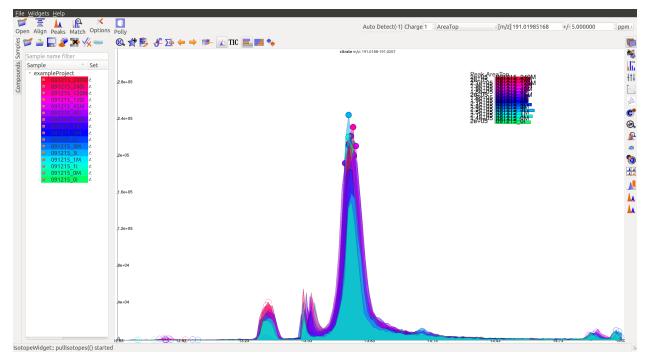
- Gaussian shape
- Perfect grouping, narrow retention time
- Good sample intensities
- Low blank intensities
- QCs look good
- An observable trend in intensity bars of standards, as well as samples.

Some examples of bad peaks are given below:

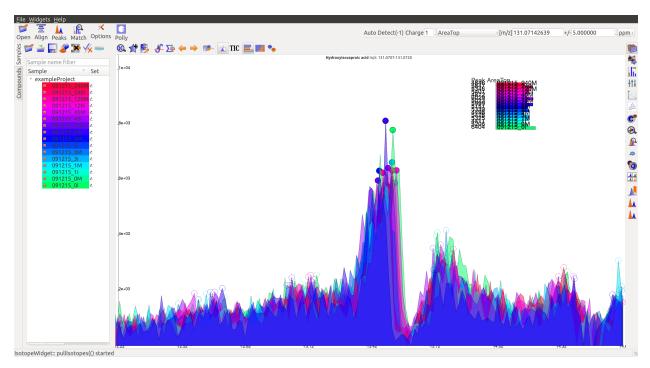
• The peaks do not have a Gaussian shape. Low intensity peaks are not grouping well. QC intensities (10⁴) are too high with respect to the low sample intensities (10²), which are very close to the noise level.



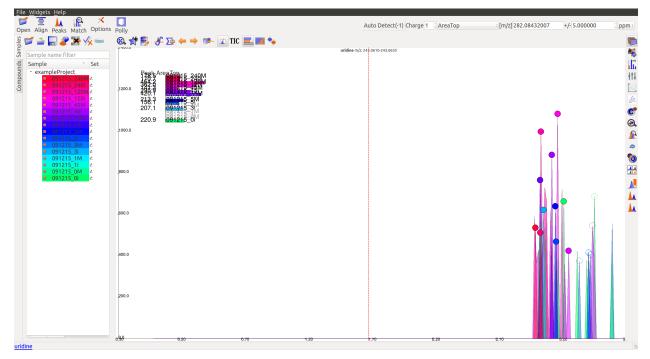
• The peaks have a good Gaussian shape. But the blank intensity bars are high. All the sample intensity bars are shorter or roughly equal to the blank intensities, implying that the peaks are noisy. This should be marked bad if better groups of the same metabolite are available.



• The intensity levels are high. The blank intensities are lower. However, the peaks are spread over a long range of retention time, have poor grouping, and have forward trailing peaks. If the signal to noise ratio was improved, this peak would probably not be detected.



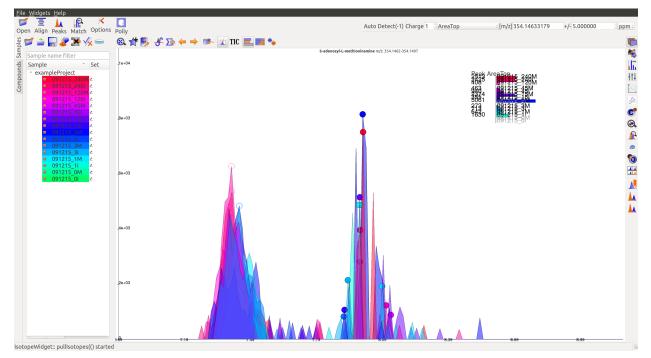
• In the following image, many sample intensities are missing from the intensities bar plots. Peaks do not have a Gaussian shape, nor good grouping. These peaks are probably noise which have been wrongly annotated. The blank intensities are high as well.



• This is a noisy group. There are no discrete peaks visible in the image. The X-axis is crowded with noise. The peak shape is sharp, triangular, or line-like; not Gaussian. The intensity levels are high, but so are the noise levels.

File	<u>W</u> idgets <u>H</u> elp				i i
۲	n Align Peaks Match Options		Auto Detect(-1) Charge 1 . AreaTop	- [m/z] 282.08432007 +/- 5.00000	0 : ppm :
Ope	en Align Peaks Match Options	Polly			
hple	🎽 🚔 🛃 🥙 🎬 🌿 🥯		oxobutanoate m/# 101.0239-101.0249		
Sar	Sample name filter	2	oxobucanoace m/2 101.0239-101.0249		
nds	Sample Set	2.1e+04			
Compounds	 exampleProject Ø91215 240M				+†+
Con	 ○ 091215 240i ○ 091215 120M 				\sim
					۵
	Ø 091215_45M Ø 091215 45i	1.8e+04			C ⁹
	■ 091215_15M ■ 091215_15M			9:58=83 89 1213 0M	
	■ 091215_5M <				
	■ 091215_5i ■ 091215_3M	-1.5e-04			
	Ø 091215_3i				
	🛛 091215_1i 🖉				
	Ø91215_0M Ø91215_0M				
		3 26-04			
		Se+03			
		.5e+03			
		3e+03			
	······	95 9.45 9.95 10.45	10.95 11.45	11:95 12:45	
2-0×	obutanoate				19

• The peaks don't have a Gaussian shape, and are also noisy. The intensity values are very low.



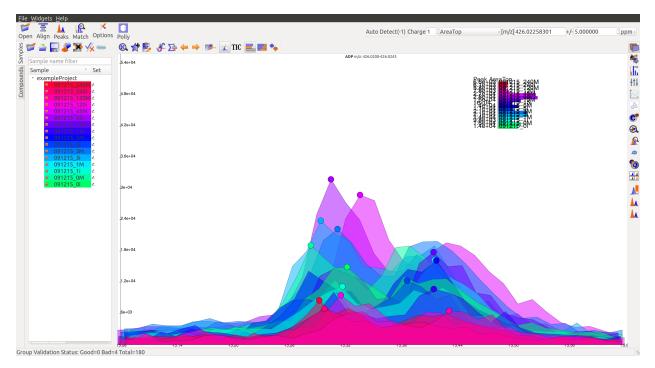
• For low intensity groups like this, the peak characteristics can be determined by zooming in.

<u>File W</u> idgets <u>H</u> elp				
Open Align Peaks Match Options		Auto Detect(-1) Charge 1 . AreaTop	: [m/z] 426.02258301	+/- 5.000000 . ppm :
Open Align Peaks Match Options	Polly	Auto Detect(=1) charge 1 . Arearop	- [11/2] 420.02258501	+/- 5.000000 . ppm :
ම 📂 🗳 🔜 🤣 🌋 🌾 👄 Eg Sample name filter	🙉 🛫 👼 🚜 🏣 👄 🔿 🃂 🗽 TIC 🔜 🎫 🍫			6
lue Cample name filter	2.7e+05	ADP m/z: 426.0200-426.0243		*\$ 11
ප Sample * Set				
* exampleProject			Poak AreaTon	illi illi
ĕ 091215_240M ≤	2.4e+05			+T+
© 091215_240i				\sim
			Pet D4	2
Ø91215_45M				69
■ 091215 15M <	.2.1e+05	9		0
Ø91215_15i				[
Ø 091215_SM ∠ Ø 091215_Si ∠				<u></u>
Ø 091215_3M	_1.8e+05			ø
Ø91215_3i Ø91215_3i Ø91215_1M				\$
				1
Ø91215_0M Ø91215_0M	.1.5e+05			
091215 <u>0</u> 12				
				4.
	.1.2e+05			
	.9e+04			
	6e+04			
	38+04			
	4.18 19.49 19.78 110.08	10.38 10.08 10.99		16.58 - 16
400				

The mouse can be used to select the area of the peak as shown below

Open Align Peaks Match Option:	s Polly	Auto Detect(-1) Charge 1 . AreaTop	: [m/z] 426.02258301 +/- 5.000000	. ppm :
ें जू 🚰 🔚 🏈 🔀 🌾 👄	8. ∰ ∰ £ ½ ← → ⊯ ⊥ TIC = = = %	ADP m/z: 426.0200-426.0243		0
경 Sample name filter 월 Sample ^ Set		AUP myz: 426.0200-426.0243		1 1 1
Sample Set • exampleProject • 091215_240M 2 • 091215_240M 2	2.40+05			+++
SOTETS_1201 ≤ SOTETS_1201 ≤				[⊻. ≫ ©
 ✓ 091215_45M ✓ 091215_45i ✓ 091215_15M 	.2.1e+05			C
 Ø 091215 15i Ø 091215 5M Ø 091215 5M Ø 091215 5i 				
Ø 091215_3M	1.8e+05			
■ 091215_1M				1
<mark>≇ 091215_0i</mark> ⊄	J.5e+05			
	1.20+05			
	1.2000			-
	_9e+04			
	.5e+04			
	3+04			
۵DP	14.16 H4.76 H4.76	13.36 13.06)	

On zooming, it will be easy to make a decision on peak quality



The user can mark any ambiguous peaks as good, and can review all such peaks later in the process.

Export

There are multiple export options available for storing marked peak data. Users can either generate a PDF report to save the EIC for every metabolite, export data for a particular group in .csv format, or export the EICs to a Json file as shown below.

File Widgets Help												
Open Align Peaks Match Option	S Polly					A	uto Detect(-1) Cha	arge 1 :AreaTop	:	[m/z] 218.10336304	+/- 5.000000	: ppm :
Sample name filter	· · · · · · · · · · · · · · · · · · ·	🔶 🔶 📂 🔝 TIC	= = *									C
Sample name filter					pa	intothenate m/z:	218.1023-218.1045					* \$.[[]
성 Sample * Set	.4.8e+06								Peak Ares	TOBIE DAMA		nie
Sample Set • exampleProject • 091215_2401 ¢ • 091215_2401 ¢	- T								2:BETHE			1111
e 091215_240M∠									3.48146			+++
등 <mark>◎ 091215_240i</mark> <	_4.2e+06					_			2:58:188			K
O91215_120IVI ≤									3:28168			
Ø 091215_120i Ø 091215_45M										SIS TM		2
■ 091215 45i <	,3.6e+06								2:38‡88	891213_8M		C ⁹
Ø 091215_15M												ě
■ 091215_15i <	3e+06											> C Q P
Ø91215_SM ≤ Ø91215_Si ≤	36+06											P
■ 091215 3M <												
	2.4e+06											
■ 091215_1M <												1
■ 091215_1i ■ 091215_0M												
■ 091215_0M €	1.8e+06											
	1.2e+06											
	_6e+05											
· · · · · · · · · · · · · · · · · · ·												
		10.46	10.76		11.08		11.30	11.00		11.96	12.26	
Peak Table 1 🍢 🥩 寒 💡	🌁 🤔 🍪 😵 🛪	🖻 📨										- 8
# ^ ID	Observe		rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		
33 glucose-1-phosphate	259.022		6.61	0.69	16	16	15	5.3e+04	15	0.82		
32 AMP	346.0562		15.32	4.02	16	14	39	1e+04	7	0.80		
31 AMP	346.056		11.57	0.27	16	16	59	2.1e+04	1686	0.85		
 ribose-phosphate ribose-phosphate 	229.0120 229.0120		7.55 6.84	1.55 0.84	16 15	5 9	35 15	8.7e+03 6.8e+03	970 830	0.83 0.85		
28 pantothenate	218.1034		13.66	2.66	16	9	57	1.2e+04	11	0.82		
27 pantothenate	218.1034	218.1034	11.27	0.27	16	16	21	3.8e+06	34	0.84		
26 deoxyribose-phosphate	213.0170		5.64	2.10	15	9	8	8.5e+03	1160	0.85		7
25 D-glucarate 24 tryptophan	209.0304 203.082		13.28 7.79	0.28 0.29	16 16	8 16	9 36	6.1e+03 1.5e+06	500 33	0.84 0.84		×
pantothenate	205.002.	, 200.0020	1.15	0.29	10	10	50	1.50.00	55	0.04		5

Users can select All, Good, Bad or Selected peaks to export.

File W	idgets <u>H</u> elp													
📁 Open	Align Peaks Match Option	IS Polly						A	uto Detect(-1) Ch	arge 1 . AreaTop	:	[m/z] 218.10336304	+/- 5.000000	: ppm :
Samples Sa	🍙 🔜 🥔 ጅ 🌾 😑	🙉 📌 👼	🖌 🖓 🏕 🔶	🃂 🔟 TIC	. . •	•								1
	mple name filter						P	antothenate m/z	218.1023-218.1045					*
spuno	mple [^] Set exampleProject	_4.8e+06									Peak Are			ili.
Sa Sa	 ○ 091215_240M ○ 091215_240i 	_4.2e+06												tti To /
0	 ○ 091215_120M ○ 091215_120i 										10000+1000 10000+1000			<u>○</u> .
	<mark>◎ 091215_45M</mark>	.3.6e+06						8			2:88+08			C°
	◎ 091215_15M													[≫ © ⊛. ∳
	■ 091215_5M	_3e+06												
	Ø 091215_3M Ø 091215_3i Ø 091215_3i Ø 091215_1M ℓ	,2.4e+06						6						- 0
	■ 091215_1M ≥ ■ 091215_1i ≥ ■ 091215_0M ≥													
	■ 091215_0i	_1.8e+06												
		_1.2e+06												
		-												
		.6e+05												
		i dhe OC) (0) 0 10.46 (10.78		11.08	Y/	11.30	11.00		11.96	12.28	
Peak	Table 1 🛃 😵 🔋	💌 🚀 🧞	📜 🔛 📾 📁											- ®
# •	ID		O Export Select	ed ected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		1
33	glucose-1-phosphate		25 Export All Gro		6.61	0.69	16	16	15	5.3e+04	15	0.82		
32	AMP		34 Export Good	0558	15.32	4.02	16	14	39	1e+04	7	0.80		
31 30	AMP ribose-phosphate		34 Export Bad 229.0120	0558 229.0119	11.57 7.55	0.27	16 16	16 5	59 35	2.1e+04 8.7e+03	1686 970	0.85 0.83		
29	ribose-phosphate		229.0120	229.0119	6.84	0.84	15	9	15	6.8e+03	830	0.85		
28	pantothenate		218.1034	218,1034	13.66	2.66	16	9	57	1.2e+04	11	0.82		
27	pantothenate		218.1034	218.1034	11.27	0.27	16	16	21	3.8e+06	34	0.84		
26	deoxyribose-phosphate		213.0170	213.0170	5.64	2.10	15	9	8	8.5e+03	1160	0.85		0
25	D-glucarate		209.0304	209.0303	13.28	0.28	16	8	9	6.1e+03	500	0.84		e
24	tryptophan		203.0825	203.0826	7.79	0.29	16	16	36	1.5e+06	33	0.84		



The *Export Groups to CSV* option lets the users save the 'good'/'bad' labels along with the peak table. Users also have the option to filter out rows that have a certain label while exporting the table.

Generate PDF Report option

Export EICs to Json option



saves all EICs with their corresponding bar plots in a PDF file.



exports all EICs to a Json file.

Another option is to export the peak data in .mzroll format that can be directly loaded into El-MAVEN by clicking on the Load *Samples*|*Projects*|*Peaks* option in the File menu. For this, go to the File option in the menu bar, and click on 'Save Project'.

<u>File W</u> idgets <u>H</u> elp													
Load Samples Project Load Classification M							А	uto Detect(-1) Cha	arge 1 . AreaTop		[m/z] 218.10336304	+/- 5.000000	: ppm :
Load Compound List Save Project As	Ctda	s 🎝 🖻 🕹 🤸	тс 👘		-								C
Save Settings	Cut				•	pa	antothenate m/z	: 218.1023-218.1045					*
Load Settings Options		6								Reak Area	TOBLE OLON		
Add S3 credentials													+†+
Report Bugs! Exit	Ctrl+	5 6								12201100			In the
✓ 091215 ✓ 091215	20M 2						8			13:88+06 43:48+06	M		
© 091215 © 091215	15M e	+06					20			1000-1000			
■ 091215_*	5M 2									2.00100	001210_0		
Ø91215_ Ø 091215		06											© (3), }
Ø91215_ ■ 091215							12						
■ 091215 ■ 091215	Si e 24	+06					6						0
■ 091215	li e												
■ 091215_0 ■ 091215_0		+06											
	1.20	+06											
	.6e+	05											
· · · · · · · · · · · · · · · · · · ·	10/18	0 0 0 10.46 · · · · · · · · · · · · · · · · · · ·		10.78		11.08		11.30	11.00		41.96	12.26	
Peak Table 1 🛛 👔	🖉 🛠 💡 🔀 🤇	🐉 🦃 🏂 🛸 🐞											- @
# 1D		Observed m/z	Expected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		î
33 glucose-1-phos 32 AMP	phate	259.0227 346.0562	259.0225 346.0558	6.61 15.32	0.69 4.02	16 16	16 14	15 39	5.3e+04 1e+04	15 7	0.82		
31 AMP		346.0561	346.0558	11.57	0.27	16	16	59 35	2.1e+04	1686	0.85 0.83		
30 ribose-phosph 29 ribose-phosph		229.0120 229.0120	229.0119 229.0119	7.55 6.84	1.55 0.84	16 15	5 9	35 15	8.7e+03 6.8e+03	970 830	0.83		
28 pantothenate		218.1034	218.1034	13.66	2.66	16	9	57	1.2e+04	11	0.82		
27 pantothenate		218.1034	218.1034	11.27	0.27	16	16	21	3.8e+06	34	0.84		
26 deoxyribose-pl 25 D-glucarate	nosphate	213.0170 209.0304	213.0170 209.0303	5.64 13.28	2.10 0.28	15 16	9 8	8	8.5e+03 6.1e+03	1160 500	0.85 0.84		3
24 tryptophan		203.0825	203.0826	7.79	0.28	16	16	36	1.5e+06	33	0.84		-

Labeled LC-MS Workflow

This is a tutorial for processing Labeled LC/MS data files through El-MAVEN.

Preprocessing

msConvert is a command-line/GUI tool that is used to convert between various mass spectroscopy data formats, developed and maintained by proteoWizard. Raw data files obtained from mass spectrometers need to be converted to certain acceptable formats before processing in El-MAVEN.

Input

msConvert supports the following formats:

- .mzXML
- .mzML
- .RAW ThermoFisher
- .RAW Walters
- .d Agilent
- .wiff ABSciex

The settings used for msConvert as a GUI tool are captured in the following screenshots:

🖳 MSConvertGUI (64-bit)		-	٥	×
List of Files File of file names File: Add Remove	About MSConvert			
	MS Level v Levels:			
	Add Remove			
	Fiter Parameters			
Output Directory:				
Browse				
Options				
Output format: mzML v Extension: Binary encoding precision:				
Write index: Use zlib compression:				
TPP compatibility: Package in gzip:				
Use numpress linear compression:				
Use numpress short logged float compression:				
Use numpress short positive integer compression:				
Use these settings next time I start MSConvert	0		Sta	rt

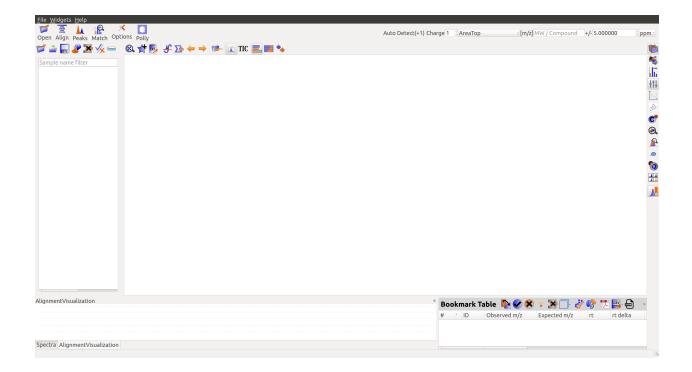
Note: zlib compression is enabled by default in msConvert. El-MAVEN in its current form does not support zlib compression. It is important to uncheck "Use zlib compression" box.

Output

msConvert can convert to an array of different formats but El-MAVEN primarily uses .mzXML and .mzML formats.

Launch EI-MAVEN

Once sample files are ready for processing, launch El-MAVEN.



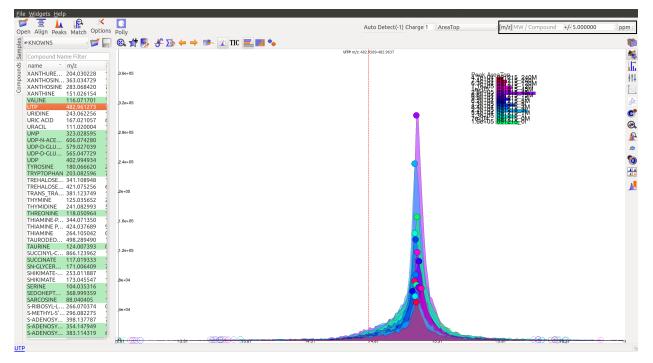
Adjust Global Settings



Global Settings can be changed from the *Options* dialog . There are 9 tabs in the dialog. Each of these tabs has parameters related to a different module in El-MAVEN. For example, a tab for Instrumentation details, a tab for File Import settings etc.

Open Align Peaks Match Options Polly	l V		Auto Detect(+1) Char	ge 1 . AreaTop	: [m/z] MW / Compound	+/- 5.000000 . ppm
	🤊 😼 🏊 👄 🔿 📂 🔝 TIC 🔜 💷 🍫					1
Sample name filter	Options Instrumentation File Import Peak Detection Peae Peak Detection and Grouping Settings EIC Smoothing Algorithm EIC Smoothing Window	ak Filtering Isotope Detectic	on EIC (XIC) [BETA] Peak Group Baseline Calculation Drop top x% intensities from		Exter •	
	Max Retention Time difference between Peaks		Baseline Smoothing	5 scans		
		Use Default Para	imeters		_	
AlignmentVisualization			8	Bookmark Table 📕	è 📀 🙁 🐑 😹 🗍 🤞	i 🖧 🎵 🔭 😵 🖗
				# [•] ID Observ		rt rt delta
Spectra AlignmentVisualization						
apecere Augrimencersuauzacion					,	

The m/z option scans the groups to find any specific m/z value and plot its corresponding EIC. The +/- option to its right is to specify the expected mass resolution error in parts per million (ppm).



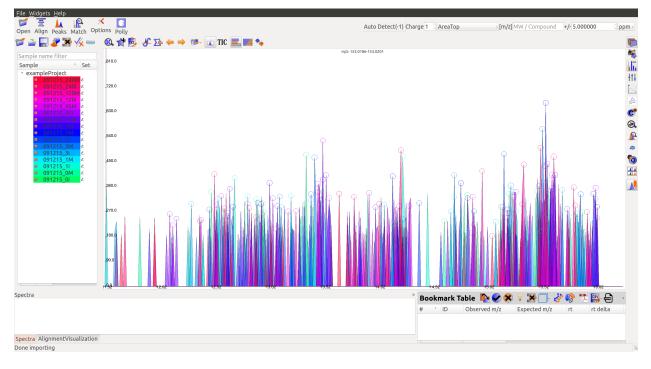
To know more about the functionality of different tabs and their settings, users can see the Widgets page. Please be sure to set the desired settings before processing an input file.

Load Samples

Users can go to File in the Menu and click on *Load Samples*|*Projects*|*Peaks* option. Then navigate to the folder containing the sample data, select all .mzXML or .mzML files and click Open. A loading bar displays the progress at the bottom.

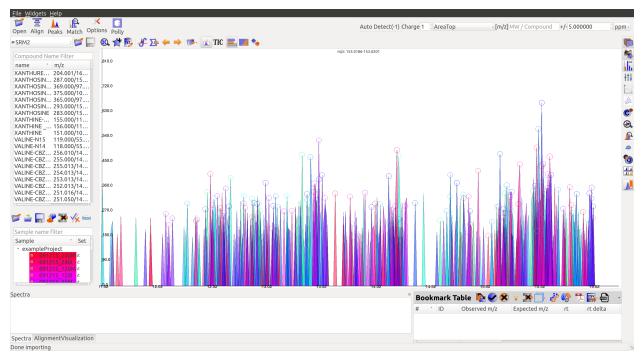
<u>File W</u> idgets <u>H</u> elp				
Load Samples Projects Peaks Load Classification Model	Ctrl+O_	Auto Detect(+1) Charge 1 AreaTop	: [m/z] MW / Compound +/- 5.000000	: ppm :
Load Compound List Save Project As	ctrl+s 😼 💤 🏣 👄 📂 🔝 TIC 📃 📰 🍫			1
Save Project As Save Settings Load Settings Options Add 53 credentials Report Bugsl Egit	Ctrl+Q			
AlignmentVisualization			🔁 🕐 😵 💗 🧱 🗍 🦑 🦃 🏃 🧱 red m/z Expected m/z rt rt de	
				1

When the samples have loaded, users should see a sample panel on the left side. If it is not displayed automatically, click on the *Show Samples Widget* button on the toolbar. El-MAVEN automatically assigns a color to every sample. Users can select/deselect any sample by clicking the checkbox on the left of the sample name.



Load Compound Database

Users can click on *Compounds* option in the leftmost menu, navigate to the folder containing the standard database file, select the appropriate .csv file and click *Open*. Alternatively, users may use any of the default files loaded on start-up.



This is a sample Compound Database:

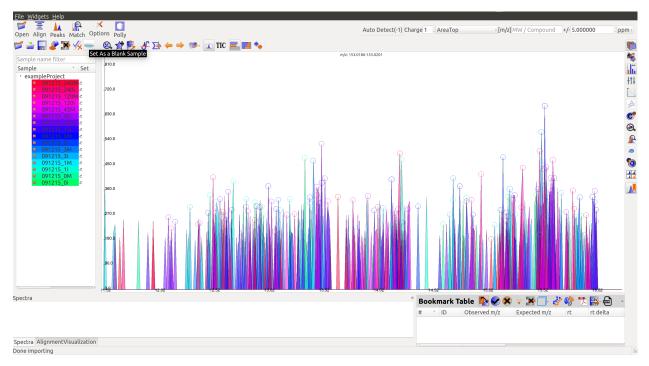
	А	В	С	D	E	F	G	н	1	J	к	L	М
1	compound	formula	id	rt	category								
2	cholesteryl sulfat	C27H46O4S	HMDB00653	14.31									
3	Deoxycholic acid	C26H43NO5	C05464	0.82									
4	trans_trans-farme	C15H28O7P2	C00448	1									
5	Cholic acid	C24H40O5	HMDB00619	13									
6	Geranyl-PP	C10H20O7P2	C00341	16.46									
7	Taurodeoxycholic	C26H45NO6S	C05463	3.86									
8	lipoate	C8H14O2S2	C00725	8.5									
9	3-hydroxy-3-met	C27H44N7O20P	C00356	1.89									
10	butyryl-CoA	C25H42N7O17P	C00630	7.45									
11	malonyl-CoA	C24H38N7O19P	C00083	6.7									
12	propionyl-CoA	C24H40N7O17P	C00100	7.7									
13	acetoacetyl-CoA	C25H40N7O18P	C00332	4.2									
14	acetyl-CoA	C23H38N7O17P	C00024	1.12									
15	dGTP	C10H16N5O13P	C00002	14.68	nucleic acids								
16	succinyl-CoA/me	C25H40N7O19P	C00091 C00683	0.7									
17	malonyl-CoA	C24H38N7O19P	C00083	11.6									
18	3-methylphenyla	C9H10O2	HMDB02222	7.35									
19	coenzyme A	C21H36N7O16P	C00010	14.14									
20	acetoacetyl-CoA	C25H40N7O18P	C00332	7									
21	Phenylpropiolic a	C9H6O2	HMDB00563	13.58									
22	3-hydroxybutyryl	C25H42N7O18P	C01144 C03561	3.5									
23	guanosine 5'-dipl	C10H11N5O17P	4	15.64									
24	Indoleactivitic acid	C11H0NIO2	HMDR00734	7 5									

It lists all metabolite names, chemical formula, HMDB ID, and the class/category of compounds they belong to (if known).

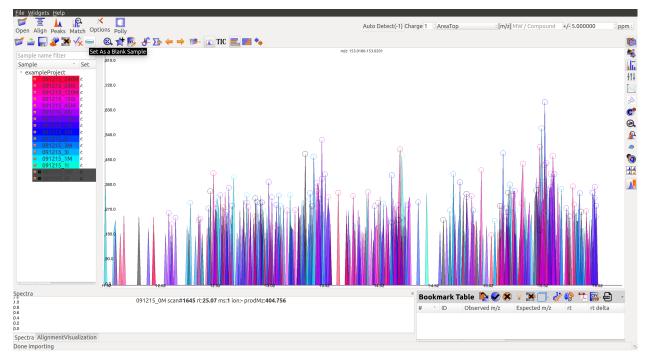
Mark Blanks

Users can mark the blanks by selecting the blank samples from the list on screen, and clicking on the Set as a Blank

Sample icon in Samples menu.



Multiple blanks can be marked together. The blanks will appear black as shown in the image below:

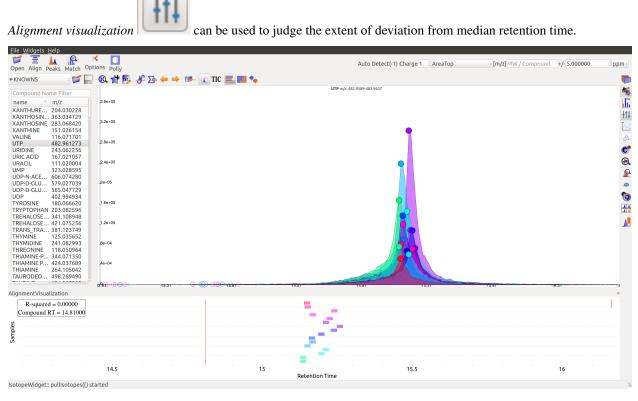


Alignment

Prolonged use of the LC column can lead to a drift in retention time across samples. Alignment shifts the peak retention time in every sample to correct for this drift and brings the peaks closer to median retention time of the group.

<u>F</u> ile <u>W</u> idgets <u>H</u> elp		
Open Align Peaks Match Options Polly	Auto Detect(-1) Charge 1 AreaTop : [m/z] MW / Compound +/- 5.000000	: ppm :
≠KNOWNS : 💅 💭 🛞 🛠 😼 🗲 🏊 👄 🃂 😰 TIC 🔜 🔜 🍫		Þ
Compound Name Filter	UTP m/z: 482.9589-482.9637	M
name * m/z 3.60+05		J.
XANTHURE 204.030228	5.281+04° 041315-240M	141
XANTHOSIN 363.034729		+†∔
XANTHOSINE 283.068420 XANTHINE 151.026154 3.20+05		\sim
VALINE 116.071701		۵
UTP 482.961273		-
URIDINE 243.062256 URIC ACID 167.021057	1968405 19684	C ? &
URACIL 111.020004		۵.
UMP 323.028595 UDP-N-ACE 606.074280 2.4e+05		R
UDP-D-GLU 579.027039		
UDP-D-GLU 565.047729		
UDP 402.994934 TYROSINE 180.066620 ^{22e+05}		6
TRYPTOPHAN 203.082596		
TREHALOSE 341.108948 TREHALOSE 421.075256		
TRANS TRA 381.123749		
THYMINE 125.035652		
THYMIDINE 241.082993 THREONINE 118.050964 J1.20+05		
THIAMINE-P 344.071350		
THIAMINE P 424.037689		
THIAMINE 264.105042 TAURODEO 498.289490		
TAURINE 124.007393		
SUCCINYL-C 866.123962 SUCCINATE 117.019333 46+04		
SUCCINATE 117.019333 4e+04 SN-GLYCER 171.006409		
SHIKIMATE 253.011887		
SHIKIMATE 173.045547		
Spectra		-
1.0 091215_0M scan#1645 rt:25.07 ms:1 ion:- prodMz:404.756	🖁 Bookmark Table 🛛 💽 😵 🔋 🔀 🗍 🤣 🦃 🏗 🔛 🖨	
0.8 0.6	# ^ ID Observed m/z Expected m/z rt rt delta	
04 02		
0.0		
Spectra AlignmentVisualization		
UTP		8

In the above image, EIC for a UTP group is displayed. If the samples were aligned, all peaks would lie at the same retention time. Since this is not the case, the samples need to be aligned.



In the above visualization, each box represents a peak from the selected group at its current retention time. Samples are said to be perfectly aligned when all peak boxes lie on the same vertical axis. The peaks are considerably scattered in the above image and therefore the samples should be aligned for better grouping of peaks.

Perform Alignment

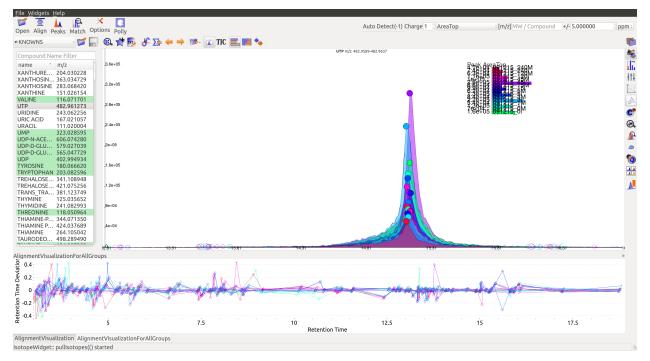
Alignment settings can be adjusted using the Align button rithm with default parameters.



. This example was aligned with Poly fit algo-

		😣 🗊 Alignment Options			
Open Align Peaks Match Opt	ions Polly	Group Selection Criteria for use in Alignment	e 1 : AreaTop	: [m/z] MW / Compound +/- 5.000000	: ppm :
	® 🛠 👼 🐇 ∑⊨ 🗢 📂 🔝 TIC				(1)
*KNOWNS	🔍 🍸 📴 🕂 🏣 🗕 💆 🔟 IIU				Þ
Compound Name Filter		Limit total number of groups in alignment to 1000 groups			N
name * m/z	3.6e+05	Peak Grouping Window 10 scans	Pook /	AreaTop	
XANTHURE 204.030228		Alignment w.r.t Expected Rt	2:481	4 6 1 215-248M	
XANTHOSIN 363.034729			List 1		+++
XANTHOSINE 283.068420	.3.2e+05	Peak Selection	6:68+1		E. I
XANTHINE 151.026154		Minumum Peak Intensity 5000.00	\$:38±0	33 881215-5M	
VALINE 116.071701		Minimum Peak S/N Ratio 2	2.481		۵
UTP 482.961273	,2.8e+05		7204		
URIDINE 243.062256 URIC ACID 167.021057		Minimum Peak Width 3 scans	1.66+0		C°
URIC ACID 167.021057 URACIL 111.020004	2.4e+05	Peak Detection Algorithm Compound Database Search :			۵,
UMP 323.028595		Select Database FKNOWNS :			
UDP-N-ACE 606.074280					R
UDP-D-GLU 579.027039	_2e+05				
UDP-D-GLU 565.047729		Alignment Algorithm			
UDP 402.994934	10.05	Alignment Algorithm Poly Fit :			0
TYROSINE 180.066620 TRYPTOPHAN 203.082596	_1.6e+05	Maximum Number of Itterations 10			
TREHALOSE 341.108948					<u></u>
TREHALOSE 421.075256	_1.2e+05	Polynomial Degree 5			
TRANS TRA 381.123749					
THYMINE 125.035652					
THYMIDINE 241.082993	Be+04				
THREONINE 118.050964					
THIAMINE-P 344.071350					
THIAMINE P 424.037689	_4e+04				
THIAMINE 264.105042 TAURODEO 498.289490					
TAURODEO 498.289490	100000 0 00 000				
	12/812/02/02/02 13:31 0 0 00 000		10.31	10.01 10.31	
AlignmentVisualization					8
R-squared = 0.00000					
Compound RT = 14.81000					
ld					
Samples					
		Status Undo Aligment			
		Status Undo Aligment			
	14.5	0% Align Cancel	5	16	
	2.71	I 0% Align Cancel	Ĭ	18	
AlianmentVisualization Alianmer	htVisualizationEorAllGroups				
IsotopeWidget:: pullisotopes(() st					1

Post-alignment the peaks in the group should appear closer to the median retention time of the group.



File Widgets Help										1
Open Align Peaks Match Op						Auto Detect(-1) Charge 1	. AreaTop	: [m/z] MW / Compound	+/- 5.000000 . pr	pm :
] 🙉 🛠 👼 J 🗵	k ← → 🌮 [🗴 TIC 📃 🗾 🍫							Þ
Compound Name Filter					UTP m/z: 482.5	9589-482.9637				
name ^ m/z	3.6e+05						F	Peak AreaTopus		eniz .
XANTHURE 204.030228							2	:48187 88 1818_248M		ш
XANTHOSIN 363.034729								(16105 881213 43M		łtļ
XANTHOSINE 283.068420	_3.2e+05									To /
XANTHINE 151.026154						•	6			Ľ.
VALINE 116.071701	2.8e+05									≽
UTP 482.961273 URIDINE 243.062256	32.00100							REDALE RELEASE AM		C
URIC ACID 167.021057										U
URACIL 111.020004	2.4e+05									۵,
UMP 323.028595										R
UDP-N-ACE 606.074280										
UDP-D-GLU 579.027039	_2e+05									۲
UDP-D-GLU 565.047729 UDP 402.994934										1
TYROSINE 180.066620	1.6e+05									
TRYPTOPHAN 203.082596										.
TREHALOSE 341.108948										
TREHALOSE 421.075256	_1.2e+05									
TRANS_TRA 381.123749										
THYMINE 125.035652 THYMIDINE 241.082993	_Be+04									
THREONINE 118.050964	300+04									
THIAMINE-P 344.071350										
THIAMINE P 424.037689	_4e+04									
THIAMINE 264.105042										
TAURODEO 498.289490							All and a second			
······································	1281-000 0		2000 3.81	14.31		14.81	15.31	15.81 00 00 000	96.31	40
AlignmentVisualization										8
R-squared = -0.88222										
Compound RT = 14.81000					-					
es										
Samples				-						
S					4					
					<u></u>					
	14.5			15		15.5			16	
				Retentio	on Time					
AlignmentVisualization Alignme	entVisualizationForAllGr	oups								
IsotopeWidget:: pullisotopes(() s	started									8

Pre-alignment, the peaks were considerably scattered while the aligned peaks lie nearly on the same axis. Users can run alignment again with different parameters if required (or with a different algorithm). Further details on Alignment settings are available on the Widgets page.

Peak Grouping

Peak grouping is an integral part of the El-MAVEN workflow that categorizes all detected peaks into groups on the basis of certain user-controlled parameters. A group score is calculated for every peak during the process. The formula for this score takes into account the difference in retention time, intensities between peaks (smaller difference leads to a better score) and any existing overlap between them (higher extent of overlap leads to better score). All three parameters have certain weights attached to them that can be controlled by the users. The formula for the score is shown in the image. More details on it can be found on the Widgets page.

(😕 🗊 Options		
	Instrumentation File Import Peak Detection Peak Grouping Score score = 1.0/((distX*A)+0.01)/ ((distY*B)+0.01)*(C*overlap) where distX = Rt difference of peaks	Peak Filtering Isotope Detection EIC (XIC) [BETA] Peak Grouping Group Rank Libraries Exter • Consider distX Weight Other 1 Overlap Weight Overlap Weight 1 	
		Use Default Parameters	

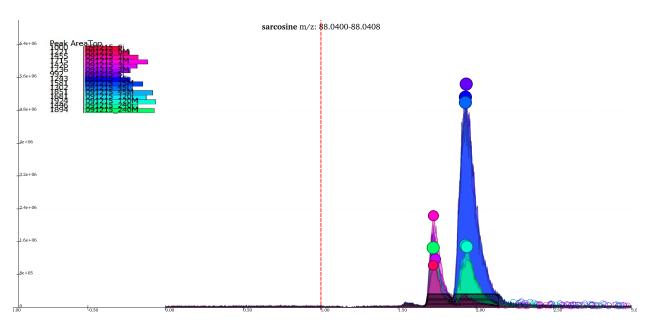
		sarcosine m/z: 88.04	400-88.0408		
_6.4e+06	Peak AreaTop 2.4e+05 AMPLE #SRK2BG7 2.2 3.9e+05 SMPLE #SRK2BG7 2.3 2.7e+05 SAMPLE #SRK2BG7 2.4 5.1e+06 SAMPLE #SRK2BG7 2.5 4.9e+06 SAMPLE #SRK2BG7 2.6 1.4e+06 SAMPLE #SRK2BG7 2.8 1.4e+06 SAMPLE #SRK2BG7 2.9 1.2e+06 SAMPLE #SRK2BG7 2.9 1.2e+06 SAMPLE				
5.6e+06	4.9e+06 SAMPL #SRK2BG7 2.6 4.4e+06 SAMPLE #SRK2BG7 2.7 1.4e+06 SAMPLE #SRK2BG7 2.8 1.4e+06 SAMPLE #SRK2BG7 2.9 1.2e+06 SAMPLE#SRK2BG7 2.10			2	
4.8e+06					
4e+06					
_3.2e+06					
_2.4e+06					
_1.6e+06				8	
8e+05					
0	I.0.50 0.00	0.50		2.00	

The above image shows two groups in the EIC window. The highlighted (solid circles) peaks belong to group A, the peaks to its left with empty circles belong to another group B. The short peaks in group A that are close to the baseline and peaks in group B come from the same samples. Additionally, the high intensity peaks of group A have a similar peak shape to group B peaks. These peaks might have been wrongly classified into separate groups because of the difference in retention time range of the two sets of peaks. The weights attached to difference in retention time and intensities, and extent of overlap can be adjusted for better grouping.



Grouping parameters can be changed from the Options dialog

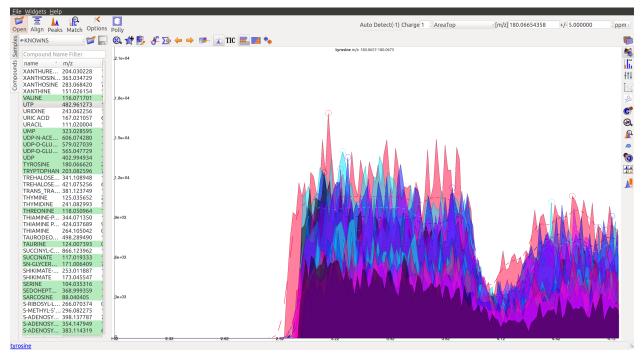
⊗	
Instrumentation File Import Peak Detection	Peak Filtering Isotope Detection EIC (XIC) [BETA] Peak Grouping Group Rank Libraries Exter
<pre>Peak Grouping Score score = 1.0/((distX*A)+0.01)/ ((distY*B)+0.01)*(C*overlap) where distX = Rt difference of peaks</pre>	 ✓ Consider distX Weight ○ 0 distY Weight ○ 0 Overlap Weight ¬○ 1
	Use Default Parameters



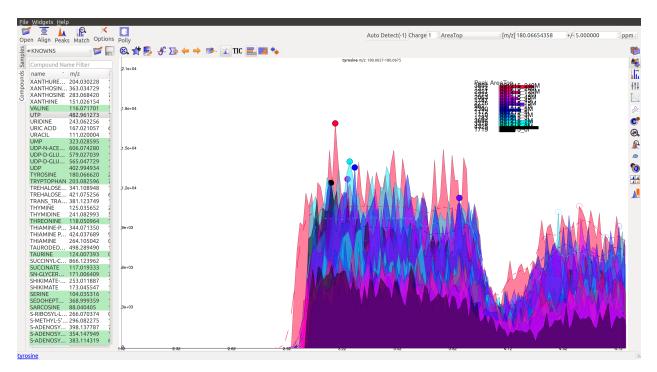
Giving less priority to difference in retention time and intensities results in the two groups being merged into a single group while the peaks that lay close to the baseline are no longer classified as valid peaks.

Baseline

When measuring a number of peaks, it is often more effective to subtract an estimated baseline from the data. This baseline should be set where ideally no peaks occur. Although sometimes the program sets a particular baseline such that one or more peaks occur below that baseline value. In the following image, the dashed line represents each baseline:



The corresponding peaks are indicated with solid circles:



The baseline correction can be done in the *Peak Detection* tab by clicking on *Options* button:

	😣 回 Options			
4	Instrumentation File Import Peak Detection Pea	k Filtering Isotope Detecti	on EIC (XIC) [BETA] Peak Grouping Group Rank	Libraries Exter
	Peak Detection and Grouping Settings		Baseline Calculation	
	EIC Smoothing Algorithm	Savitzky-Golay :	Drop top x% intensities from chromatogram	2%
1	EIC Smoothing Window	10 cans		
	Max Retention Time difference between Peaks	0.25 i min	Baseline Smoothing	5 scans
		Use Default Par	ameters	

Further details on settings can be accessed here.

Isotope Detection

Samples are labeled?

The Peaks dialog

can be used to detect labeled peaks along with the unlabeled ones in the Peaks Table.

n		Ք № 🗢 № 🔝 TIC 🔜 🎫 🍫	m/z: 482.9589-482.9637			100 militaria 100 militaria
Sample name filter Sample name filter sampleProject 001215_2400 c 001215_2400 c	.3.6e+05					
* exampleProject	-	🛛 💿 Peak Detection				+++
© 091215_240M ≪ © 091215 240i ≪	3.2e+05	Beak Detection				TI
O 091215_120M ≤ 091215_120I ≤		Feature Detection Selection Group Filtering Method	Summary			
<mark>≝ 091215_45M</mark>	.2.8e+05	Automated Feature Detection: Find peaks by sli Mass Domain Resolution (ppm) 20.000		10000000000		
Ø 091215_45i		Time Domain Resolution (scans) 10 scans	Limit m/z Range 0.00	100000000000		C
Ø 091215_15i	2.4e+05	Auto Detect And Ignore Isotopes	Limit Intensity Range 50000.00	9999999999.00		۵,
<mark>⊠ 091215_5M</mark>						R
Ø91215_3M Ø91215_3M	_2e+05	Compound Database Search: Limit slices to set o Select Database	f known m/z and retention time valu KNOWNS	es		
■ 091215_1M <	1.6e+05	EIC Extraction Window +/- PPM	5.00 ppm			0
■ 091215_1i < ■ 091215_0M <	1.00+05	Match Retention Time (+/-)	1.00 min	•		
<mark>⊯ 091215_0i</mark>	1.2e+05	Limit Number of Reported Groups per Compound	10 best	:		
	_8e+04	Match Fragmenation Precursor Match Tollerance 0.00	Report Isotopic Peaks			
		Fragment Peak Matching Tollerance 0.00	Isotope Detect	ion Options		
	_4e+04	Minumum Match Score 0.00				
AlignmentVisualization	2812/03000	Status			15.81 16.31	
Augminenterisoduzación		0%		Find Peaks Cancel		
		Use D	efault Parameters			

On opening the *Feature Detection Selection* tab, the *Report Isotopic Peaks* box must be checked. Clicking on the *Isotope Detection Options* gives the following window. Alternately, these settings can also be accessed from the *Options* dialog.

File Widgets Help	Dellu		Auto Detect(-1) Charge 1	AreaTop : [m/	z]MW / Compound +/- 5.000000	: ppm :
Sample name filter		🎍 👄 📂 🔝 TIC 📃 🧱 🍫				C
Sample name filter Sample ^ Set • exampleProject	_3.6e+05		m/z: 482.9589-482.9637			*
© 091215_246M ∉ © 091215_240i ∉ © 091215_120M ∉	_3.2e+05		Q			+t+ [
● 091215_1201	_2.8e+05					✓
■ 091215_15i	2.4e+05 💿 Option					<u>ک</u>
<mark>⊯ 091215_3M</mark>	Samples Ar	tion File Import Peak Detection Peak Filtering Isotope Del re Labeled?	Filter Isotopic Peaks based on these crite			- 0
Ø 091215_1i	1.6e+05 Bookmark	k, peakdetection & file export ■ D2 ■ C13 ■ N15 ■ S34	Isotope is within [X] scans of parent	5 scans :		
	Isotopic w	vidget	Maximum %Error to Natural Abundance	100.00 %		
	.4e+04	Use Default			_	
					-	
AlignmentVisualization	12:81.7 (0) (0) (0)	13.31 O (0) USO 13.810 14.31	14.81 11	5.31 15.81	18.31	
Deers interaction						

Done importing

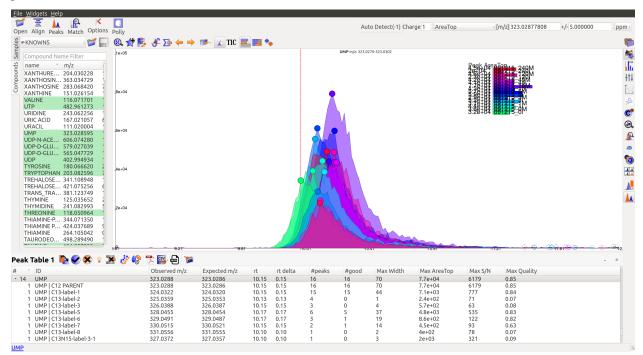
- *Bookmarks, peak detection, file export*: To select the labeled atoms that should be used in bookmarking, peak detection and export. D2: Deuterium, C13: Labeled carbon, N15: Labeled nitrogen, S34: Labeled sulphur.
- *Isotopic widget*: To select the labeled atoms that should be displayed in the isotopic widget. D2: Deuterium, C13: Labeled carbon, N15: Labeled nitrogen, S34: Labeled sulphur.
- Number of M+n isotopes: To set the maximum number of labeled atoms per ion in the experiment.

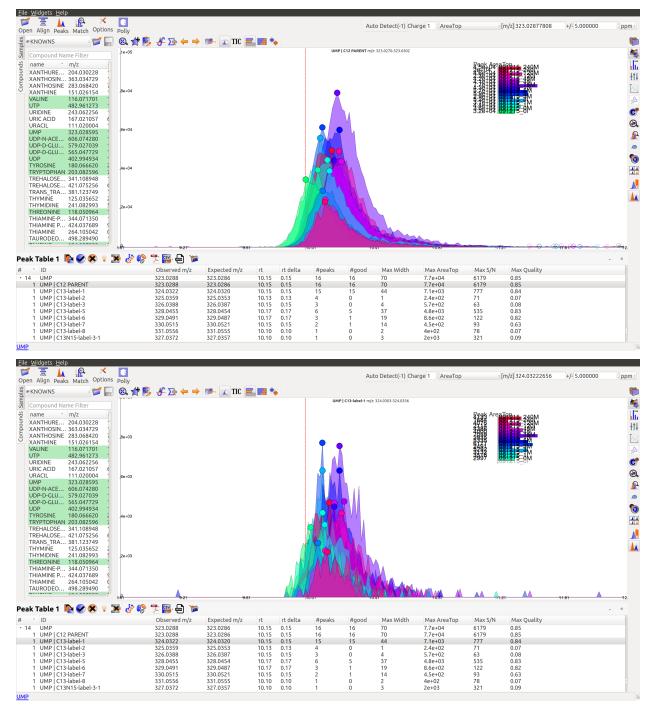
• *Abundance Threshold*: To set the minimum threshold for isotopic abundance. Isotopic abundance is the ratio of intensity of isotopic peak over the parent peak.

Filter Isotopic Peaks based on these criteria

- *Minimum Isotope-Parent Correlation*: To set the minimum threshold for isotope-parent peak correlation. This correlation is a measure of how often they appear together.
- *Isotope is within [X] scans of parent*: To set the maximum scan difference between isotopic and parent peaks. This is a measure of how closely they appear together on the retention time scale.
- *Maximum % Error to Natural Abundance*: To set the maximum natural abundance error expected. Natural abundance of an isotope is the expected ratio of amount of isotope over the amount of parent molecule in nature. Error is the difference between observed and natural abundance as a fraction of natural abundance.
- Correct for Natural C13 Isotope Abundance: The box should be checked to correct for natural C13 abundance.

In the image below, Peak Table 3 has a drop down button with metabolites that shows all labeled isotopologues of that particular metabolite.

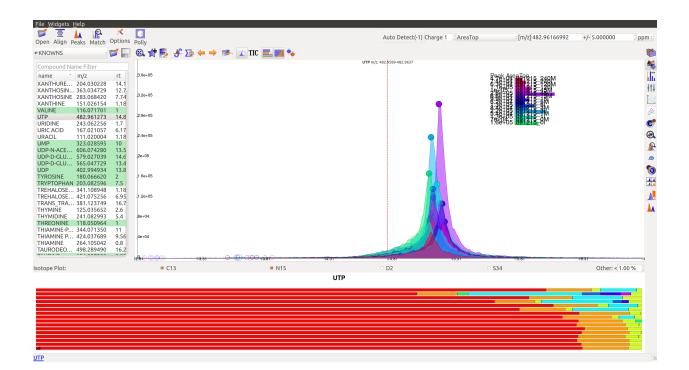




Show Isotope Plots



This icon **a source of the selected group is shown** below. Each bar in the plot represents the relative percentage of different isotopic species for the selected group in a sample.



Mass Spectra



Mass Spectra Widget displays each peak, its mass, and intensity for a scan. As the widget shows all detected masses in a scan, the ppm window for the EIC and consequently grouping can be adjusted accordingly. This feature is especially useful for MS/MS data and isotopic detection.

<u>File W</u> idgets <u>H</u>							· · · · · · · · · · · · · · · · · · ·	
	🔺 🔐 🗡					Auto Detect(-1) Charge 1 AreaTop	= [m/z] MW / Compound +/- 5.000000	: ppm :
	aks Match Options							
Se KNOWNS Compound N	: 📁 🗖) 🙉 📌 👼	, 🔏 🎠 🚧 🏓	🕨 🃂 🦹 TIC 🔜 📑	•			1
du						TP m/z: 482.9589-482.9637		1
Compound N								
	* m/z	_3.6e+05					4.724-04-07-0215 240M	. III
S XANTHURE.	204.030228						6:38184 881218=128M	+++
E XANTHOSIN	363.034729 E 283.068420	3.2e+05						11+
XANTHOSIN	151.026154						8:28:84 84218-35	\sim
VALINE	116.071701					The second se		۵
UTP	482.961273	.2.8e+05						
URIDINE	243.062256						1.6e+05 091215 01**	C
URIC ACID URACIL	167.021057 (111.020004	2.4e+05						
URACIL	323.028595	1.40105						
	606.074280							
UDP-D-GLU.	579.027039	2e+05						
	565.047729 1							
UDP	402.994934	1.6e+05						0
TYROSINE	180.066620 2 N 203.082596 7	1.00100						
	341.108948							
TREHALOSE	421.075256 (.1.2e+05						A
	381.123749							_
THYMINE	125.035652 2							
THYMIDINE	241.082993 5 118.050964	8e+04						
	344.071350							
THIAMINE P.	424.037689	4e+04						
THIAMINE	264.105042 (
TAURODEO.	498.289490							
		12980-0000-	13:31	- @@@@@@ Ga.ar	14.31	19.81 10.31	10.81 030 030631 0	
Spectra								8
3.6e+05				091215_4	45i scan# 998 rt: 15.19 ms: 1 ion:	- prodMz: 482.961		
3.2e+05				482.961395 ^{z=1}				
2.8e+05								
2.4e+05								
2e+05								
1.6e+05								
1.2e+05								
8e+04								
.4e+04					483.964752			
8 0.96	481.66	481.977600	482.36	483.06	483.76	484.46 .484.965759	485.86485.972137 486.56	
IsotopeWidget:: p	ullisotopes(() starte	d						17

Peak Curation

Generally there are two broad workflows to curate peaks in El-MAVEN:

- Manual Peak Curation using Compound DB widget
- Automated Peak Curation

Although, for labeled data Automatic Peak Curation is not meaningful because it will not curate any labeled groups.

Manual Peak Curation using Compound DB widget

Clicking the Peaks icon

on the top opens the settings dialog.

1.6e+05							
	😣 🗉 Peak Detection						
.2e+05							
	Feature Detection Selection Group	Filtering Method S	ummary				
2.8e+05					100000000000		
1.4e+05			6				
	 Auto Detect And ignore isotop 	85	Limic incensicy Range	50000.00	. 99999999999.00 .		
te+05		nit slices to set of		ntion time value	s		
.6e+05							
.2e+05	Linic Number of Reported Group	s per compound	To Desc				
le+04	Match Fragmenation Precursor Match Tollerance	0.00	Report Isotop	ic Peaks			
	Fragment Peak Matching Tollerar	ce 0.00	1	Isotope Detectio	n Options		
le+04	Minumum Match Score	0.00					
8100000	Status					10.81 10.31	
	1	0%			Find Peaks Cancel		
		Use De	fault Parameters				
2 2 1 1	8e+05 4e+05 9+05 6e+05 2e+05	Peature Detection Selection Group as-os Aduomated Feature Detection: Mass Domain Resolution (crans) a-uto Detect And Ignore Isotop >-05 Compound Database Search: Lin Select Database as-os EIC Extraction Window +/- PPM Match Retention Time (+/-) 2x-os Limit Number of Reported Group PAtch Fragmenation Precursor Match Tollerance Fragment Peak Matching Tolleran Minumum Match Score	Feature Detection Selection Group Filtering Method S aevo5 Bit C Extraction Window +/- PPM Match Resolution (aevo) Match Retention Time (+/-) Limit Number of Reported Groups per Compound aevo4 avo4 aevo5 aevo4 aevo5 aevo5 <td>Feature Detection Selection Group Filtering Method Summary ae-05 Advomated Feature Detection: Find peaks by sticing m/z and retention Mass Domain Resolution (pm) 20.000 Himt Time Domain Resolution (scans) To scans Imme Domain Resolution (scans) To scans - Auto Detect And Ignore Isotopes - Auto Detect And Ignore Isotopes - Compound Database Search: Limit Slices to set of known m/z and retention Select Database Be-05 EIC Extraction Window +/- PPM S.00 ppm - Match Retention Time (+/-) Limit Number of Reported Groups per Compound I best Be-04 Minumum Match Score 0.00 Hinumum Match Score 0.00</td> <td>Feature Detection Selection Group Filtering Method Summary aevids Aith Retention Time (r/) Limit Number of Reported Groups per Compound Image Tragment Peak Matching Tollerance Precursor Match Tollerance Aith Fragment Peak Matching Tollerance Aith Match Score Aith Status Aith Status</td> <td>Feature Detection Selection Group Filtering Method Summary ae.o5 Automated Feature Detection: Find peaks by slicing m/z and retention time space Mass Domain Resolution (ppm) 4e.o5 • Auto Detect And Ignore Isotopes • Compound Database Search: Limit Slices to set of Known m/z and retention time values Select Database ee.o5 EIC Extraction Window +/- PPM Solo ppm Imit Number of Reported Groups per Compound 100 best Imit Number of Reported Groups per Compound Precursor Match Tollerance Precursor Match Tollerance 0.00 Status 1 0%</td> <td>Baseds Feature Detection Selection Group Filtering Method Summary Baseds Automated Feature Detection: Find peaks by slicing m/z and retention time space Mass Domain Resolution (ppm) 20.000 1 mme Domain Resolution (scans) 10 scans 0* Auto Detect And Ignore Isotopes Limit Inte Range 0.00 • Auto Detect And Ignore Isotopes Limit Inte Range 0.00 • Auto Detect And Ignore Isotopes Limit Intensity Range 5000.00 • Compound Database Search: Limit slices to set of known m/z and retention time values Select Database Select Database KNOWNS Imit Number of Reported Groups per Compound 10 best 1 meter Detection Options Match Fragmenation •Report Isotopic Peaks isotope Detection Options Procursor Match Tollerance 0.00 0.00 isotope Detection Options isotope Detection Options Procursor Match Tollerance 0.00 0.00 isotope Detection Options isotope Procursor Match Tollerance 0.00 0.00 isotope Detection Options isotope Procursor Match Tollerance 0.00 0.00 isotope Detection Options isotope Procursor Match Tollerance 0.00 0</td>	Feature Detection Selection Group Filtering Method Summary ae-05 Advomated Feature Detection: Find peaks by sticing m/z and retention Mass Domain Resolution (pm) 20.000 Himt Time Domain Resolution (scans) To scans Imme Domain Resolution (scans) To scans - Auto Detect And Ignore Isotopes - Auto Detect And Ignore Isotopes - Compound Database Search: Limit Slices to set of known m/z and retention Select Database Be-05 EIC Extraction Window +/- PPM S.00 ppm - Match Retention Time (+/-) Limit Number of Reported Groups per Compound I best Be-04 Minumum Match Score 0.00 Hinumum Match Score 0.00	Feature Detection Selection Group Filtering Method Summary aevids Aith Retention Time (r/) Limit Number of Reported Groups per Compound Image Tragment Peak Matching Tollerance Precursor Match Tollerance Aith Fragment Peak Matching Tollerance Aith Match Score Aith Status Aith Status	Feature Detection Selection Group Filtering Method Summary ae.o5 Automated Feature Detection: Find peaks by slicing m/z and retention time space Mass Domain Resolution (ppm) 4e.o5 • Auto Detect And Ignore Isotopes • Compound Database Search: Limit Slices to set of Known m/z and retention time values Select Database ee.o5 EIC Extraction Window +/- PPM Solo ppm Imit Number of Reported Groups per Compound 100 best Imit Number of Reported Groups per Compound Precursor Match Tollerance Precursor Match Tollerance 0.00 Status 1 0%	Baseds Feature Detection Selection Group Filtering Method Summary Baseds Automated Feature Detection: Find peaks by slicing m/z and retention time space Mass Domain Resolution (ppm) 20.000 1 mme Domain Resolution (scans) 10 scans 0* Auto Detect And Ignore Isotopes Limit Inte Range 0.00 • Auto Detect And Ignore Isotopes Limit Inte Range 0.00 • Auto Detect And Ignore Isotopes Limit Intensity Range 5000.00 • Compound Database Search: Limit slices to set of known m/z and retention time values Select Database Select Database KNOWNS Imit Number of Reported Groups per Compound 10 best 1 meter Detection Options Match Fragmenation •Report Isotopic Peaks isotope Detection Options Procursor Match Tollerance 0.00 0.00 isotope Detection Options isotope Detection Options Procursor Match Tollerance 0.00 0.00 isotope Detection Options isotope Procursor Match Tollerance 0.00 0.00 isotope Detection Options isotope Procursor Match Tollerance 0.00 0.00 isotope Detection Options isotope Procursor Match Tollerance 0.00 0

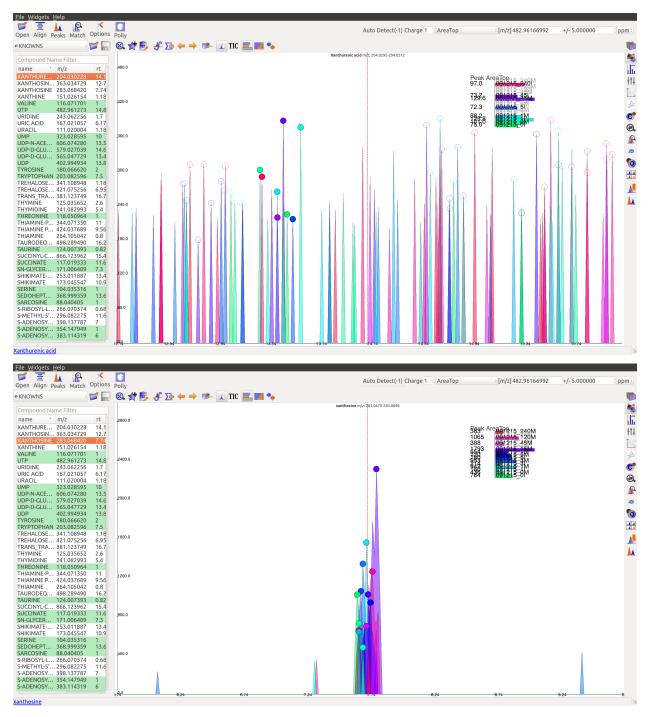
Users must check the box for Report Isotopic Peaks in the Group Filtering tab.

Note: Users should not click on Find Peaks after checking the box for manual curation. Clicking on that option would

start Automatic Peak Detection. For adjusting other settings, users can access them through the Options dialog .

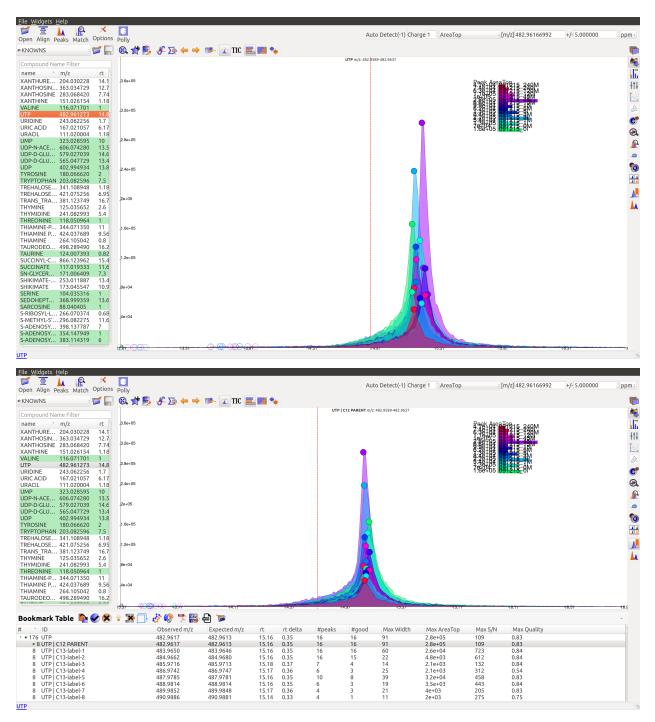
For more details on how to access Peak Detection settings, read this Widgets page.

To use manual curation using the compound DB widget, users have to iterate over all the compounds in the compound DB on the extreme left of the window, as highlighted in the images below.



Once on a compound, El-MAVEN shows the highest ranked group for that m/z. Users can now choose a group or reject it. There are two ways to do this.

In the first workflow, users need to double click on the peak group of his choice. This will get the retention time line to the median of the group and also add the metabolite to the bookmarks table (as shown in the image below). Users can read more about the bookmarks table here.



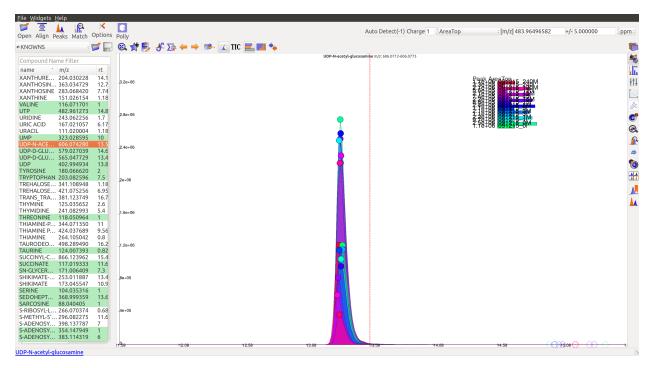
When the users select the first group they would be asked if they would like to auto-save the state of the application. This feature allows the users to go back to their curated peaks if they so wishes in future.

File Widgets	<u>H</u> elp													
1 =		×							A.,	to Dotoct(1) C	harge 1 . AreaTop	: [m/z] 482.96166992	+/- 5.000000	: ppm :
Open Align (Peaks Match	Options	Polly						Au	to Detect(-1) C	marge i . Arearop	· [III/2] 402.90100992	+/- 3.000000	. ppm-
■KNOWNS	:	1	🙉 쑭 👼	🗲 🎦 🍕		TIC 📕	L 🖬 🍫							r
Compound Na	ame Filter								UTP m/z: 482.95	89-482.9637				🧠
	m/z	rt	.3.6e+05											1.
XANTHURE XANTHOSIN		14.1												+t+
XANTHOSIN.		12.7 7.74	3.2e+05											
XANTHINE	151.026154	1.18									•			\sim
VALINE UTP	116.071701	1 14.8	_2.8e+05											۵
URIDINE	482.961273 243.062256	14.8	1.00.00											C ?
URIC ACID	167.021057	6.17												ě
URACIL UMP	111.020004 323.028595	1.18	.2.4e+05											
UDP-N-ACE		10 13.5												[
UDP-D-GLU	579.027039	14.6	2e+05											
UDP-D-GLU UDP	565.047729 402.994934	13.4 13.8												1
TYROSINE	180.066620	2	1.6e+05					😣 Autosave						
TRYPTOPHAN	203.082596	7.5						Do you wa	nt to enable					
TREHALOSE TREHALOSE		1.18 6.95	1.2e+05					autosave?						
TRANS TRA		16.7	5.20105						No Yes					
THYMINE	125.035652	2.6								J /				
THYMIDINE	241.082993	5.4 1	_8e+04											
THIAMINE-P		11								11				
THIAMINE P		9.56	_4e+04											
THIAMINE TAURODEO		0.8 16.2								dist of				
TAORODEO	450.205450	10.2	1200000	113	a - O	തന്തം	0	14.31				10.01	10.31	11
Bookmark	Table 陷			🤌 🧞 🎵	🛐 📾	1								-
# ^ ID	Observed n		Expected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality			
↑ 1 UTP	482.9617		482.9613	15.19	0.38	16	16	91	2.8e+05	109	0.83			
Group Validatio	n Status: Good	d=0 Bad=	0 Total=1											8

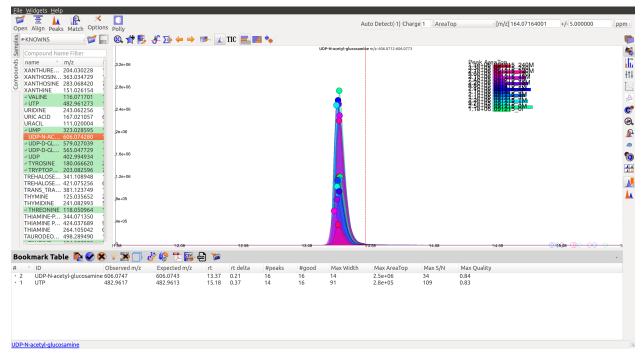
<u>File Widgets H</u>elp Open Align Peaks Match Options Polly Auto Detect(-1) Charge 1 . AreaTop : [m/z] 483.96496582 +/- 5.000000 ppm : KNOWNS 📁 🔜 🞕 💏 😼 🗲 🏊 👄 🔿 📂 📠 TIC 🔜 🎫 🍫 P 3.6e+04 UTP | C13-label-1 m/z: 483.9622-483.9670 ***** name m/z XANTHURE... 204.030228 XANTHOSIN... 363.034729 XANTHOSINE 283.068420 XANTHINE 151.026154 rt 14.1 12.7 7.74 1.18 3.2e+04 łtł \sim 2 80+04 VALINE UTP URIDINE URIC ACID URACIL 》 C Q A A A 116.071701 482.961273 243.062256 167.021057 111.020004 323.028595 606.074280 570.074280 1 14.8 1.7 6.17 1.18 UMP UDP-N-ACE. +04 13.5 14.6 13.4 13.8 2 7.5 1.18 6.95 16.7 2.6 5.4 1 UDP-N-ACE... 606.074280 UDP-D-CLU... 579.027039 UDP-D-CLU... 579.027039 UDP UDP 180.066620 TRYPTOPHAN 203.082596 TRPHALOSE... 421.075256 TRANS_TRA... 381.123749 THYMINE 125.035652 THYMIDINE 241.082993 THREDNINE 118.050964 THAMINEPA... 344.071350 **1** .6e+04 1.2e+04 Be+03 THIAMINE-P... 344.071350 THIAMINE P... 424.037689 THIAMINE 264.105042 TAURODEO... 498.289490 11 9.56 0.8 16.2 103 Bookmark Table 🐚 🥪 🛠 🕴 🍱 🗍 🤣 🎨 🏗 🔤 🐞
 BOOKMARK IADLE
 Image: Constraint of the second Expected m/z 482.9613 483.9646 484.9680 485.9713 486.9747 487.9781 488.9814 489.9848 490.9881 rt delta 0.35 0.35 0.35 0.35 0.37 0.36 0.35 0.35 0.36 0.33 Observed m/z 482.9617 483.9650 484.9662 485.9716 486.9742 487.9785 488.9814 #peaks Max Width Мах АгеаТор Max S/N Max Quality #good rt 15.16 15.16 15.16 15.18 15.17 15.16 15.16 15.16 15.17 15.14 2.8e+05 2.8e+05 2.6e+04 4.8e+03 2.1e+03 2.1e+03 3.2e+04 3.5e+03 4e+03 2e+03 109 109 723 612 132 312 458 443 205 275 0.83 0.83 0.84 0.84 0.84 0.54 0.83 0.83 0.83 0.75 16 16 16 7 6 10 91 91 22 14 25 39 19 21 11 16 16 16 15 4 488,9814 6 4 4 489.9852 490.9886 3 UTP

Users can then use dropdown arrow on bookmarked group to mark all its isotopologues as good or bad.

After marking all the groups for a compound, users can scroll down to the next compound and decide on the basis of shown EIC if the group should be marked for curation.



Double clicking on any peak (solid coloured circle) moves the retention time line along the group. And the group moves to the bookmark table.



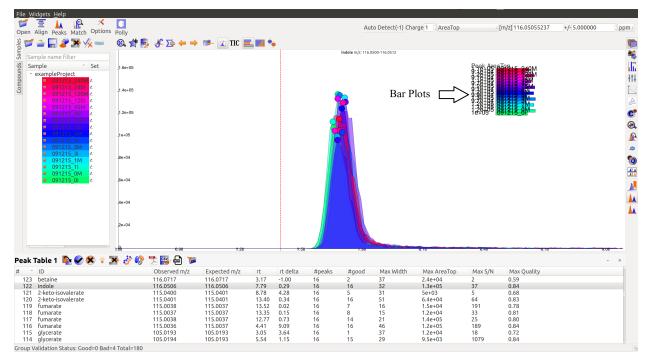
Qualifying peaks as good or bad is explained in the next section.

Guidelines for Peak Picking

• A peak's width and shape are two very crucial things to look at while classifying a peak as good or bad. The peak's shape should have a Gaussian distribution and width should not be spread across a wide range of retention

time.													
<u>F</u> ile <u>W</u> idgets <u>H</u> elp									-				
📁 🗄 🛕 😤 🗡													
Open Align Peaks Match Options	s Polly						Au	uto Detect(-1) Cha	rge 1 : AreaTop	:	[m/z] 116.05055237	+/- 5.000000	: ppm :
		<u>-</u> <u>∽⊾</u> 🥧 🛶	🥦 🛣 TIC 📕	2 📷 🍕									1
E		<u></u>		`			indole m/z: 116	.0500-116.0512					
Sample ^ Set	_1.6e+05									Last area	19215 248M		
Sample Set • exampleProject • 091215_240M • 091215_240M • 091215_240M										9:18:18			+†∔
S 091215 240i <	1.4e+05									1:48:183			In 1
	1.40100					<u> </u>				9:38:83			<u> </u>
◎ 091215_120i ◎ 091215_45M							P			1:48:18			۵
<mark>፼ 091215_45i</mark> ∉	_1.2e+05									12405 ⁵	91213_8M		C ⁹
Ø 091215_15M							21						© © & &
■ 091215_15i ■ 091215_5M	1e+05												
Ø 091215_5i	16403												<u></u>
■ 091215_3M <								Gaussian I	Distibution				۵
■ 091215_3i ■ 091215_1M	_8e+04							Ouussiun L	isticution				0
Ø 091215 1i													S
Ø 091215_0M Ø 091215_0I	_6e+04												
[∞] 091215_01 [∞]	_D0+04												
	_4e+04												
	2e+04												
	20+04												
	3.88	6.98	7.28		7.58		V.85		0.10	0.40	0.78	9.	
Peak Table 1 🍢 🔗 🔋	🐱 🚀 🥵 💌	🚳 📾 🌉					/idth	-1					- 8
# ^ ID		served m/z	Expected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		
123 betaine		5.0717	116.0717	3.17	-1.00	16	2	37	2.4e+04	2	0.59		
122 indole	116	5.0506	116.0506	7.79	0.29	16	16	32	1.3e+05	37	0.84		
121 2-keto-isovalerate		.0400	115.0401	8.78	4.28	16	5	31	5e+03	5	0.68		
120 2-keto-isovalerate		5.0401 5.0038	115.0401 115.0037	13.40 13.52	0.34 0.02	16	16 7	51 16	6.4e+04 1.5e+04	64	0.83 0.78		
119 fumarate 118 fumarate		5.0038	115.0037	13.52	0.02	16 16	8	15	1.2e+04	191 33	0.81		
117 fumarate		5.0037	115.0037	12.77	0.73	16	14	21	1.4e+05	25	0.80		
116 fumarate		5.0036	115.0037	4.41	9.09	16	16	46	1.2e+05	189	0.80		
115 glycerate		5.0193	105.0193	3.05	3.64	16	1	37	1.2e+03	18	0.72		
114 glycerate		5.0193	105.0193	5.54	1.15	16	15	29	9.5e+03	1079	0.84		
Group Validation Status: Good=0 Bad													
Group valuation status, GOOd=0 Bad	-4 10001-100												1.0

• Peak Intensities for a group are plotted as bar plots for all the samples. These bar plots have heights relative to the other samples. Thus, for a good peak the intensities should be high.



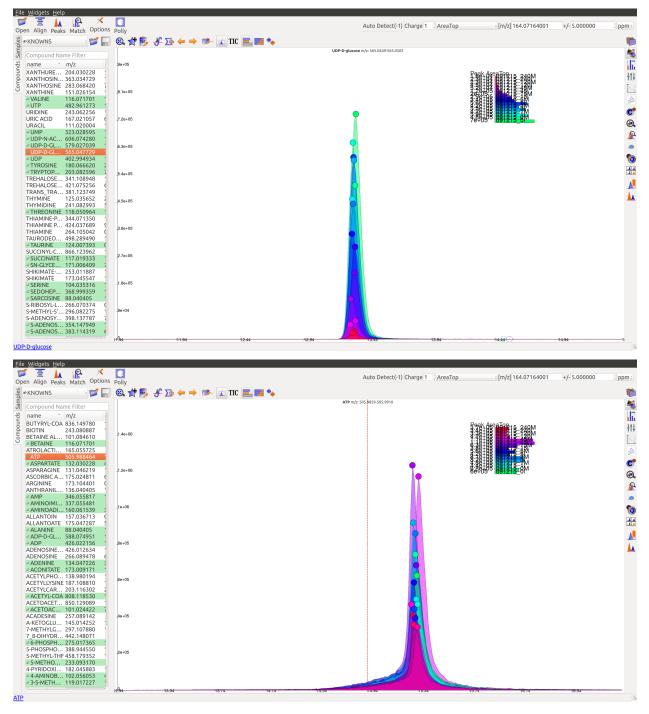
- Intensity Barplot heights should be higher for all the samples than Blank samples, as shown above. We use intensities of Blank samples to set our group baseline. Blank intensities are used to calibrate intensity values across zero concentration.
- A good peak should have standards with varying intensity in a particular fashion (increasing or decreasing).
- Quality Control (QC) samples give us information about the quality of the data, i.e., it assesses reproducibility

time.

and software performance. Samples whose intensities and concentrations are already known are used as QCs to determine if the instrument are working as expected. Values (and scales) can be calibrated using QCs.

- If peak groups of a particular metabolite are separated apart (not aligned well) then we should use stringent alignment parameters to overcome this problem.
- For a particular metabolite, let's say if it has n number of groups, then the group which is much closer to the above guidelines should be selected as good peak. Multiple groups can also be selected in case of ambiguity (if retention time information is not provided).

A good peak would look similar to the following peaks:

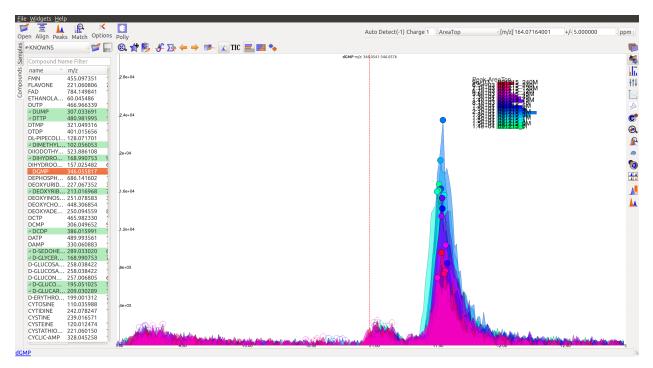


File Widgets He	lp										
🔰 🗐 🛔	P 🗶										
	ks Match Options	Polly					Auto Detect(-1) Charg	je 1 . AreaTop	: [m/z] 164.07164001	+/- 5.000000	: ppm :
je reknowns	: 📁 📊	🙉 党 🌄	 y 🖓 🆛 🗲 🍅	🕤 主 TIC 📃 📕	📕 🎭						6
See KNOWNS Compound N		1				NADH m/z	664 1142-664.1208				*
	ame Filter										
NICOTINATE NICOTINAM NICOTINAM	m/z										
NICOTINATE	122.024750	2.4e+05							Reak AreaTop		141
NICOTINAM.	. 333.049316								3.787.04 83.213-248		+†+
									4 38194 881318 12W		To /
	. 201.135696								3:28+84 83 213 73M		<u> </u>
NADPH	744.083862 1								2:58±83 881212 5M		۵
NADP+	742.068176 1	2.1e+05							1:38:483 881213 3M		-
NADH	664.117493 1								7:88+84 88 8 8 8		C
NAD+ NAD	662.101868 8								5e+04 ⁴ 091215_01		C ° Q
N-CARBAM N-ACETYLP											~
N-ACETYL-L		1.8e+05									R
	. 130.050964 7										
	187.072433										
	. 188.056442										0
	. 300.048981 7										S
N-ACETYL	300.048981	1.5e+05									
	179.056107 🏾 🏅										
METHYLNIC											
METHYLM											
METHYLCYS.											
METHIONIN.		_1.2e+05									
- METHIONI											
MALONYL-C.											
LYSINE	133.014252 1 145.098251 (
LIPOATE	205.036255	9e+04									
LEUCINE/IS											
✓ LACTATE	89.024422 (
L-ARGININO											
KYNURENINE	207.077515										
KYNURENIC.		6e+04									
	191.019730 1										
	267.073486										
	186.056046										
INDOLE-3-C.		22.04									
✓ INDOLE IMP	116.050575 7 347.039825	_3e+04									
IMIDAZOLE											
IDP	427.006165										
	135.031235										
C C		12914	12.54	12.04	19.94	274				-	
NADH		14.14	16.07	16.07	10.04	0.14	0.47.14	14.04		· ·····	15.

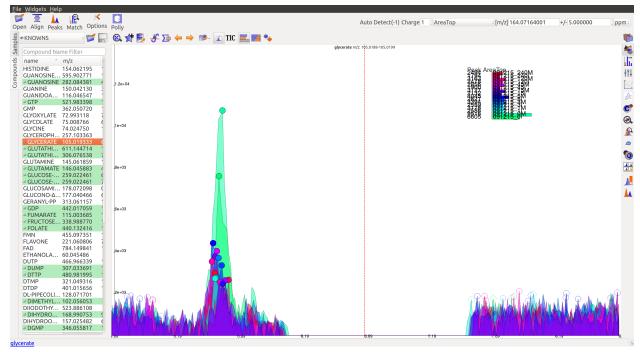
- Gaussian shape
- Perfect grouping, narrow retention time
- Good sample intensities
- Low blank intensities
- QCs look good
- An observable trend in intensity bars of standards, as well as samples.

Some examples of bad peaks are given below:

• The peaks have a good Gaussian shape. But the blank intensity bars are high. All the sample intensity bars are shorter or roughly equal to the blank intensities, implying the peaks are most likely noise.



• The intensity levels are low relatively. The peaks are spread over a long range of retention time. They have poor shape, poor grouping and lie close to noise. If the signal to noise ratio was improved, this peak would probably not be detected.



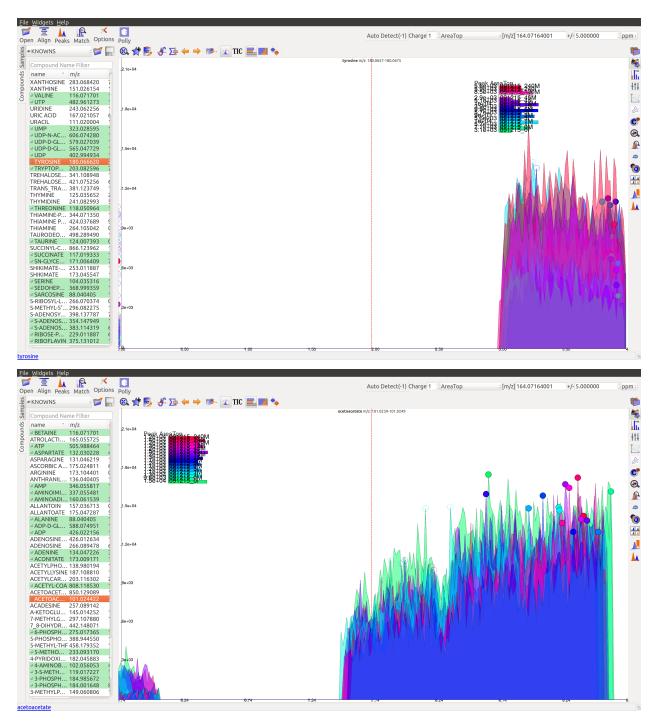
• In the following image, many sample intensities are missing from the intensities bar plots. Peaks do not have a Gaussian shape, nor good grouping. These peaks are probably noise which have been wrongly annotated.

File	Widgets Help												
1	E II	P. ×											
		s Match Options						Auto Detect(-1) Charge	1 AreaTop		: [m/z] 164.07164001	+/- 5.000000	: ppm :
			-										
a e	KNOWNS	: 📁 💭	🛞 🛟 🗖	🍹 🗲 🎦 🔶	🃂 🖈 TIC 💻	L 📰 🎭							
Ê	KNOWNS Compound Na			· · · · ·			N-carbamovi-L-asp	rtate m/z: 175.0352-175.0369					
Sa	Compound Na	me Filter											1
		m/z	3600.0										
ŝ	NG-DIMETH									Pook A	ooTop		
8	NADPH	744.083862								184	eaTop 91215_240		+†↓
Compounds	NADP+	742.068176 1									A91215 120		To /
0	NADH	664.117493 1	3200.0							776	891215-15M		<u> </u>
	≤ NAD+	662.101868 8								3350	881318-5M		<u></u>
	N-CARBAM									778923 29232 159292 2005	091215 31		
	N-ACETYLP									2022	181313 1M		C
	N-ACETYL-L N-ACETYL		,2800.0							2065	091213_0I**		C * @
	N-ACETYL		.2000.0					T T					
	N-ACETYL												
	N-ACETYL												
	N-ACETYL		2400.0					M					
	MYO-INOS		2400.0										0
	METHYLNIC												
	METHYLM												
	METHYLCYS												
	METHIONIN METHIONI		2000.0										
	MALONYL-C												
	MALATE	133.014252											
	YSINE	145.098251 (
	LIPOATE	205.036255	_1600.0										
	LEUCINE/IS												
	LACTATE	89.024422 6											
	L-ARGININO												
	KYNURENINE		1200.0										
	VINURENIC												
	NOSINE	267.073486											
	INDOLEAC												
	NDOLE-3-C		800.0										
	INDOLE	116.050575 7											
	MP	347.039825											
	MIDAZOLE												
	DP	427.006165	400.0										
	HYPOXANT												
	HYDROXY HYDROXY								4			1	
	HYDROXYP				1				A A			4	
2	~ ~ ~		998	10.75		6	1.75	19.95				19.75	
N-car	hamovl-L-asoa	ortato	10.20	10.75	911.2	J 1		HELES.	16.19		10.60	10.70	14

• This is a noisy group. There are no discrete peaks visible in the image. The X-axis is crowded with noise. The peak shape is sharp, triangular, or line-like; not Gaussian. The intensity levels are high, but so are noise levels.

\overrightarrow{e} \overrightarrow{e} \overrightarrow{e}		Auto Detect(-1) Charge 1 AreaTop	: [m/z] 164.07164001	+/- 5.000000
Open Align Peaks Match Options		Auto Detect(+1) charge 1 . Arearop	. [11/2] 104.07104001	
S KNOWNS S Compound Name Filter	🙉 🛠 👼 🗲 🏊 👄 🧇 🍽 🗽 TIC 🔜 🎫 🍫			le l
Compound Name Filter		2-oxobutanoate m/# 101.0239-101.0249		A
sp name ^ m/z ADP-D-GL 588.074951	.2.1e+04			li
은 ADP 426.022156 1				+1
ADENOSINE 426.012634 ADENOSINE 266.089478				<u>~</u>
ADENINE 134.047226 3 ACONITATE 173.009171		138		2
ACETYLPHO 138.980194	_1.8e+04	128		C
ACETYLLYSINE 187.108810 ACETYLCAR 203.116302		7:58	184 891213_8M	۲
« ACETYL-COA 808.118530 1				5
ACETOACET 850.129089 - ACETOAC 101.024422	a berok 🙀 IN			
ACADESINE 257.089142 A-KETOGLU 145.014252				1
7-METHYLG 297.107880				
7_8-DIHYDR 442.148071 				.[ال.]
5-PHOSPHO 388.944550 5-METHYL-THF 458.179352				
5-METHO 233.093170 4-PYRIDOXI 182.045883				L
4-AMINOB 102.056053				
3-S-METH 119.017227 3-PHOSPH 184.985672				
3-PHOSPH 184.001648 8 3-METHYLP 149.060806				
3-HYDROXY 852.144714 1				
3-HYDROXY 910.150208 1 2-OXOBUT 101.024422				
2-OXO-4-ME 147.012146 - 2-KETO-IS 115.040070				
2-ISOPROPY 175.061203				
2-HYDROXY 147.029892 2-DEHYDRO 193.035385				
2-AMINO 158.118652 2 3-DIPHOS 264.951996				
2_3-DIHYDR 153.019333				
I-METHYL 280.105133 1-METHYL-H 168.077850				
1_3-DIPHOP 264.951996	10 9 49 9 95 10 45	10.95 11.45	11.95	19.45
	2.50 0.50 10.40	10.00 11.40	11100	14.00

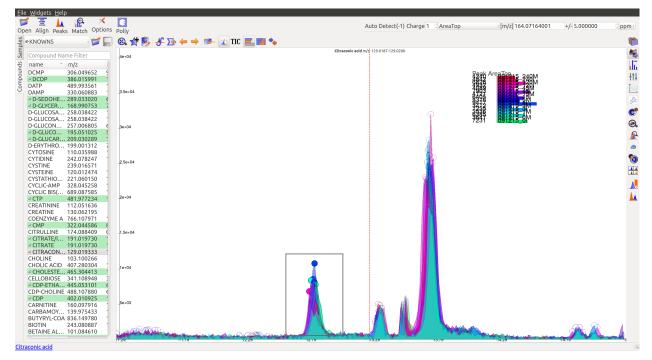
More examples of noisy peaks:



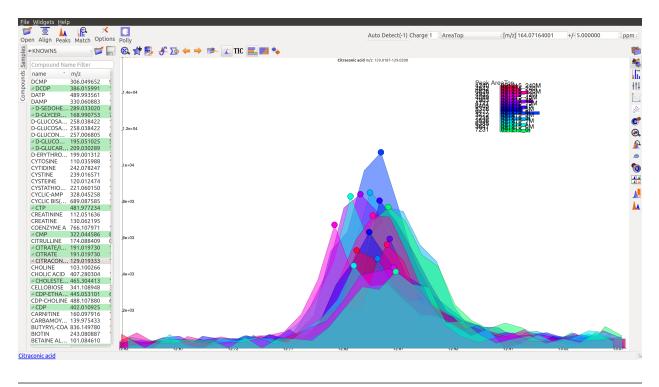
- The peaks don't have a Gaussian shape, and are also noisy. The intensity values are very low.
- For low intensity groups like this, the peak characteristics can be determined by zooming in.

File Widgets He	lp					
🗹 🚍 🚺						
	ks Match Options		Auto Detect(-1) Charge 1 . AreaTop	: [m/z] 164.07164001	+/- 5.000000	ppm :
Sald KNOWNS Compound N	: 📁 🕞	🙉 🛠 👼 🚜 🏊 🗢 🖈 😰 TIC 🔜 🎫 🍫				0
Ê			Citraconic acid m/z 129.0187-129.0200			
S Compound N	ame Filter	4e+04				*
	m/z					di.
DCMP DCDP DATP DATP	306.049652		Peak	AreaTop		
DCDP	386.015991		2829	2 10 10 10 10 10 10 10 10 10 10 10 10 10		+++
5 DATP	489,993561					T. 1
DAMP	330.060883	3.5e+04	299	881318 48		\sim
	289.033020 🤅		7917 449 449 449 449 449 449 449 449 449 44			-
	. 168.990753 7		9933	181212 - M		-
	258.038422 1		· · · · · · · · · · · · · · · · · · ·	: 881213 IM		C
	258.038422		7231	891213_8		
	257.006805	3e+04				C° @.
	195.051025 5 209.030289					P
	. 199.001312		<u> </u>			
CYTOSINE	110.035988					
CYTIDINE	242.078247					6
CYSTINE	239.016571	2.5e+04	M			
CYSTEINE	120.012474					
CYSTATHIO	. 221.060150					
CYCLIC-AMP	328.045258					
	. 689.087585					
CTP	481.977234 1	20+04				
	112.051636					
CREATINE	130.062195					
	766.107971					
	322.044586 8					
	. 191.019730	_1.5e+04				
✓ CITRATE/I	191.019730 1					
	129.019333 1					
CHOLINE	103,100266					
	407.280304	10+04				
	465.304413 1	Detot				
	341.108948					
CDP-ETHA.	445.053101 (
	E 488.107880 (
CDP	402.010925 1	50+03				
CARNITINE	160.097916	peros				
	. 139.975433					
	A 836.149780					
BIOTIN	243.080887				- 🙈 🖄 - I	4
DE TAINE AL.	. 101.084610	A CONTRACT AND A CONTRACT	Barbychard April 20 19 19 19 19	Martin Martin Martin	and the second	Δ.
		Tiza 11.7a 12.2a 12.7a	II3.29 II3.79	14.29	14.79	-11
Citraconic acid						8

The mouse can be used to select the area of the peak as shown below



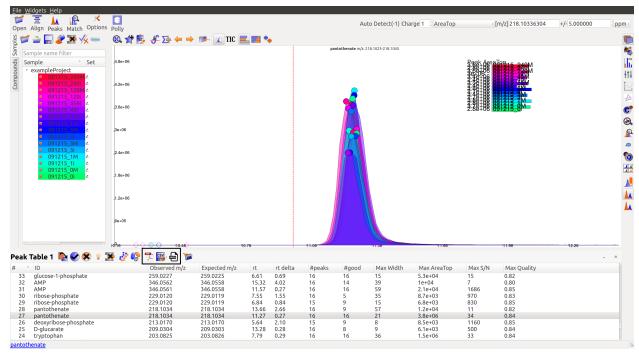
On zooming, it will be easy to make a decision on peak quality



Note: Users can mark any ambiguous peaks as good, and can review all such peaks later in the process.

Export

There are multiple export options available for storing marked peak data. Users can either generate a PDF report to save the EIC for every metabolite, export data for a particular group in .csv format, or export the EICs to a Json file as shown below.



\Xi 🗼 🔐 🗡 Align Peaks Match Option	IS Polly						Au	to Detect(-1) Cha	arge 1 :AreaTop	:	[m/z] 218.10336304	+/- 5.000000	; pp
1 🚔 🔛 🥙 🅦 🌾 💳		🦻 🗲 泽 🗕	🕨 📂 🔝 TIC 🛛	E. 🖬 🍕	•								
ample name filter						pa	ntothenate m/z: 2	218.1023-218.1045					
Sample * Set	_4.8e+06									Seak Area	19815 240M		
 exampleProject 										368+06			
 ○ 091215_240M ○ 091215_240i 	4.2e+06												
091215_120M							2			3:28108			
Ø 091215_120i Ø 091215_45M										3.68+86	SIZ 3 TM		
Ø 091215_45i	.3.6e+06									2:38‡88	91213_0 ^M		
Ø 091215_15M Ø 091215_15M													
© 091215_151 2	3e+06												
■ 091215_5i <													
Ø 091215_3M ≤ Ø 091215_3i ≤													
■ 091215_1M	,2.4e+06						121						
■ 091215_1i <													
Ø 091215_0M Ø 091215_0i	1.8e+06												
	_1.2e+06												
	_6e+05												
	i dhe O	⊖ <u>@ 0</u> <u>10,46</u>		10.78		11.08		11.30	11.00		11:96	12.28	
Table 1 🖌 🖌 🏌	ጅ 🚀 🧞												-
 ID glucose-1-phosphate 		O Export Select		rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		
glucose-1-phosphate AMP		25 Export All G		6.61 15.32	0.69 4.02	16 16	16 14	15 39	5.3e+04 1e+04	15 7	0.82 0.80		
AMP		34 Export Bad	0558	11.57	0.27	16	16	59	2.1e+04	1686	0.85		
ribose-phosphate		229.0120	229.0119	7.55	1.55	16	5	35	8.7e+03	970	0.83		
ribose-phosphate pantothenate		229.0120 218.1034	229.0119 218.1034	6.84 13.66	0.84 2.66	15 16	9 9	15 57	6.8e+03 1.2e+04	830 11	0.85 0.82		
pantothenate		218.1034	218.1034	11.27	0.27	16	16	21	3.8e+06	34	0.82		
deoxyribose-phosphate		213.0170	213.0170	5.64	2.10	15	9	8	8.5e+03	1160	0.85		
D-glucarate tryptophan		209.0304 203.0825	209.0303 203.0826	13.28 7.79	0.28	16 16	8 16	9 36	6.1e+03 1.5e+06	500 33	0.84 0.84		
		203.0823	203.0820	1.19	0.29	10	10	J0	1.50+00	22	0.04		

Users can select All, Good, Bad or Selected peaks to export.



The Export Groups to CSV option lets the users save the 'good'/'bad' labels along with the peak table. Users also have the option to filter out rows that have a certain label while exporting the table.

Generate PDF Report option



saves all EICs with their corresponding bar plots in a PDF file.

Export EICs to Json option



exports all EICs to a Json file.

Another option is to export the peak data in .mzroll format that can be directly loaded into El-MAVEN by clicking on the Load Samples|Projects|Peaks option in the File menu. For this, go to the File option in the menu bar, and click on 'Save Project'.

	idgets <u>H</u> elp													
	Samples Projects Peaks Classification Model	Ctrl+O						A	Auto Detect(-1) Cha	arge 1 : AreaTop		[m/z] 218.10336304	+/- 5.000000	: ppm :
	Compound List		}	THE THE										-
	Project As Settings	Ctri+S	* * 🖛 🗕 🗕			•	pi	antothenate m/;	z: 218.1023-218.1045					100 m
Load	Settings	2										-		
Optio Add S	ns 3 credentials	B												
Repo	rt Bugs!										2:48+86			+†∔
Exit	≝_091215_120M ≤	Ctrl+Q ⁶												\sim
	Ø 091215_120i Ø 091215_45M										3:48+98 3:88±88			۵
	Ø 091215_45i	_3.6e+06									2:38+08	891213_8M		C
	Ø91215_15M ∉ Ø91215_15i ∉													© @.
	■ 091215_5M ■ 091215_5i	.3e+06												P
	■ 091215_3M													
	■ 091215_3i ■ 091215_1M	2.4e+06												0
	■ 091215_1i													
	■ 091215_0M ≥ ■ 091215_0i <	_1.8e+06												
		1.2e+06												
		6e+05												
	-)		0 00 1040		0.76		11.00					11.90		
Peak -	Table 1 🔖 ✔ 🛠 💡	🛛 🙀 🧦 🔊	🏃 🔛 🖨 🐞											- 9
# .			Observed m/z	Expected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		•
33	glucose-1-phosphate		259.0227	259.0225	6.61	0.69	16	16	15	5.3e+04	15	0.82		
32 31	AMP AMP		346.0562 346.0561	346.0558 346.0558	15.32 11.57	4.02 0.27	16 16	14 16	39 59	1e+04 2.1e+04	7 1686	0.80 0.85		
30	ribose-phosphate		229.0120	229.0119	7.55	1.55	16	5	35	8.7e+03	970	0.83		
29 28	ribose-phosphate pantothenate		229.0120 218.1034	229.0119 218.1034	6.84 13.66	0.84 2.66	15 16	9 9	15 57	6.8e+03 1.2e+04	830 11	0.85 0.82		
27 26	pantothenate deoxyribose-phosphate		218.1034 213.0170	218.1034 213.0170	11.27 5.64	0.27	16 15	16 9	21 8	3.8e+06 8.5e+03	34 1160	0.84		
25	D-glucarate		209.0304	209.0303	13.28	0.28	16	8	9	6.1e+03	500	0.84		8
24	tryptophan		203.0825	203.0826	7.79	0.29	16	16	36	1.5e+06	33	0.84		

Labeled LC-MS/MS Workflow

This is a tutorial for processing LC/MSMS data files through El-MAVEN.

Preprocessing

msConvert is a command-line/GUI tool that is used to convert between various mass spectroscopy data formats, developed and maintained by proteoWizard. Raw data files obtained from mass spectrometers need to be converted to certain acceptable formats before processing in El-MAVEN.

Input

msConvert supports the following formats:

- .mzXML
- .mzML
- .RAW ThermoFisher
- .RAW Walters
- .d Agilent
- .wiff ABSciex

The settings used for msConvert as a GUI tool are captured in the following screenshot:

🖳 MSConvertGUI (64-bit)		-	٥	×
List of Files File of file names File: Add Remove	About MSConvert			
	MS Level Levels: 1 -			
	Add Remove			
	Filter Parameters			
Output Directory:				
Browse				
Options				
Output format: mzML v Extension: Binary encoding precision:				
Write index: Use zlib compression:				
TPP compatibility: Package in gzip:				
Use numpress linear compression:				
Use numpress short logged float compression:				
Use numpress short positive integer compression:				
Use these settings next time I start MSConvert			Sta	rt

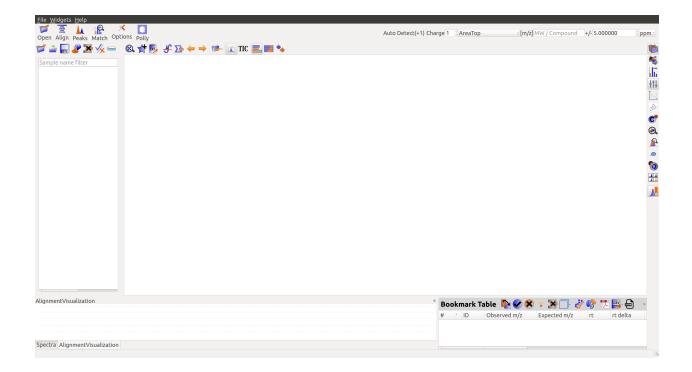
Note: It is important that zlib compression is enabled by default in msConvert. El-MAVEN in its current form does not support zlib compression. Make sure to uncheck "Use zlib compression" box.

Output

msConvert can convert to an array of different formats but El-MAVEN primarily uses .mzXML and .mzML formats.

Launch EI-MAVEN

Once sample files are ready for processing, launch El-MAVEN.



Adjust Global Settings



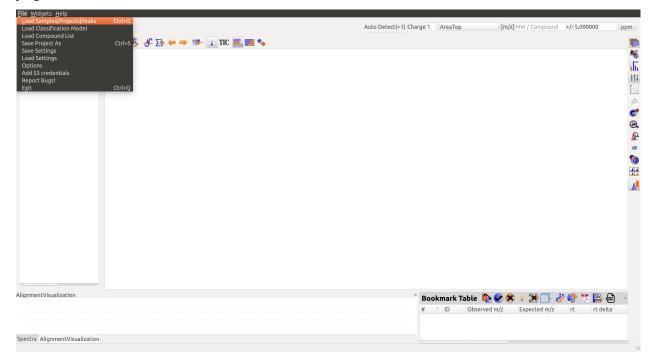
Global Settings can be changed from the *Options* dialog . There are 9 tabs in the dialog. Each of these tabs has parameters related to a different module in El-MAVEN. For example, a tab for Instrumentation details, a tab for file import settings, etc.

Sample name filter Sample name filter Simple name filter <th>Open Align Peaks Match Options Polly</th> <th>,</th> <th></th> <th>Auto Detect(+1) Char</th> <th>ge 1 . AreaTop</th> <th>: [m/z] MW / Compour</th> <th>id +/- 5.000000 : p</th>	Open Align Peaks Match Options Polly	,		Auto Detect(+1) Char	ge 1 . AreaTop	: [m/z] MW / Compour	id +/- 5.000000 : p
Sample name filter Sample name filter Sample name filter Sample name filter Sample name filter Sample name filter Sample name filter Sample name filter Sample name filter Sample name filter Sample name filter S							
Max Retention Time difference between Peaks 0.25 in min Use Default Parameters Use Default Parameters Use Default Parameters	Sample name filter	Options Instrumentation File Import, Peak Detection Peae Peak Detection and Grouping Settings EIC Smoothing Algorithm	ak Filtering Isotope Detectic Savitzky-Golay :	Baseline Calculation		s Exter + +	
LlignmentVisualization Bookmark Table No S V M I A Constraint of the second sec		Max Retention Time difference between Peaks		-	5 scans		
# * ID Observed m/z Expected m/z rt rt delta			Use Default Para	meters		_	
# * ID Observed m/z Expected m/z rt rt delta	AlignmentVisualization			2	Bookmark Table	🖪 📀 🕱 💡 💌 🗍	🤣 🧐 🎵 🔤 🖨
Spectra Alleement/Austration							
	Coactor Alicement/Jaurinetics						

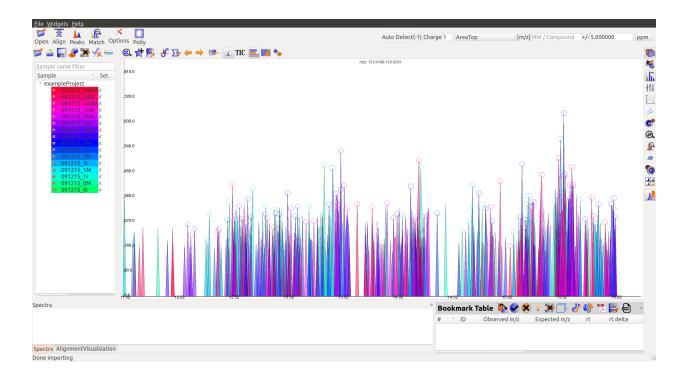
To know more about the functionality of different tabs and their settings, users can see the Widgets page. Please be sure to set the desired settings before processing an input file.

Load Samples

Users can go to File in the Menu, click on *Load Samples*|*Projects*|*Peaks* option in the File menu. Then navigate to the folder containing the sample data, select all .mzXML or .mzml files and click Open. A loading bar displays the progress at the bottom.

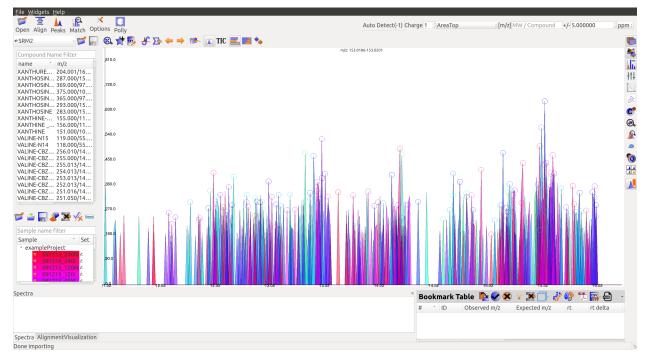


When the samples have loaded, users should see a sample panel on the left side. If it is not displayed automatically, click on the *Show Samples Widget* button on the toolbar. El-MAVEN automatically assigns a color to every sample. Users can select/deselect any sample by clicking the checkbox on the left of the sample name.



Load Compound Database

Users can click on *Compounds* option in the leftmost menu, navigate to the folder containing the standard database file, select the appropriate .csv file and click *Open*. Alternatively, users may use any of the default files loaded on start-up.



This is a sample Compound Database:

_	A	В	C	D	E	F	G	Н	l I	
	polarity	compound	precursorMz	productMz	expectedRt					
2	-	Citrate 191/67	191	67	4.55					
3	-	Citrate 192/67	192	67	4.55					
4	-	Citrate 192/68	192	68	4.55					
5	-	Citrate 193/67	193	67	4.55					
6	-	Citrate 193/68	193	68	4.55					
7	-	Citrate 193/69	193	69	4.55					
8	-	Citrate 194/68	194	68	4.55					
9	-	Citrate 194/69	194	69	4.55					
10	-	Citrate 194/70	194	70	4.55					
11	-	Citrate 195/69	195	69	4.55					
12	-	Citrate 195/70	195	70	4.55					
13	-	Citrate 195/71	195	71	4.55					
14	-	Citrate 196/70	196	70	4.55					
15	-	Citrate 196/71	196	71	4.55					
16	-	Citrate 197/71	197	71	4.55					
17	-	Citrate 191/111	191	111	4.55					
18	-	Citrate 192/111	192	111	4.55					
19	-	Citrate 192/112	192	112	4.55					
20	-	Citrata 103/112	103	112	1 55					

It lists all metabolite names, chemical formula, HMDB ID, and the class/category of compounds they belong to (if known).

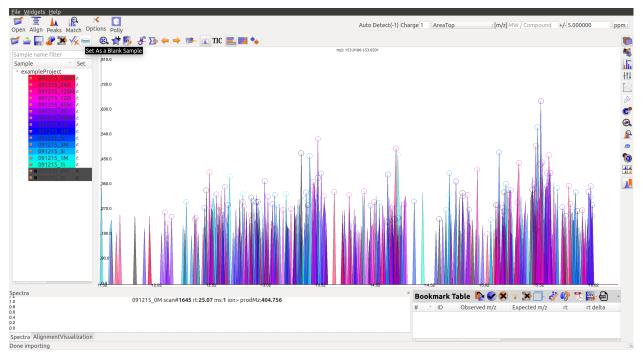
Mark Blanks

Users can mark the blanks by selecting the blank samples from the list on screen, and clicking on the Set as a Blank

Sample icon

in Samples menu.





Multiple blanks can be marked together. The blanks will appear black as shown in the image below:

Alignment

(missing)

Export

There are multiple export options available for storing marked peak data. Users can either generate a PDF report to save the EIC for every metabolite, export data for a particular group in .csv format, or export the EICs to a Json file as shown below.

File V	<u>W</u> idgets <u>H</u> elp													
1	Align Peaks Match Option	s Polly						A	uto Detect(-1) Cha	arge 1 . AreaTop	:	[m/z] 218.10336304	+/- 5.000000	: ppm :
al 📖	1 🖆 🔜 🥔 🔀 🌾 🥯	🙉 쑭 👼	🌡 🕂 🏊 🔶 🦊	🕨 📂 և TIC 🚪	2 🖬 🍬	•								C
Samples	ample name filter						pi	antothenate m/z:	218.1023-218.1045) *
θS	ample * Set	4.8e+06									Reak Area	19815 040M		. The
	 exampleProject 													łtł
Co	 ○ 091215_240i ○ 091215_120M 	_4.2e+06												\sim
	Sector = 091215 120i ≤													
	✓ 091215_45M ✓ 091215 45i	3.6e+06									2.88+08	8 2 3 8M		C
	091215_15M											_		ě
	Ø 091215_15i	3e+06												
	091215_5i													R
	Ø 091215_3M Ø 091215_3i													
	091215_1M	2.4e+06						121						0
	Ø 091215_1i													
	■ 091215_0i <	1.8e+06												
		1.2e+06												
		_6e+05												
		200400												
		id ^o le O	<u> </u>	н	0.76		11.08		11.30	11.66		11.96	12.26	
Peak	: Table 1 🖺 🖌 🛠 💡	💌 🛷 🚷	🖾 🔛 🏷											- 8
	· ID		Observed m/z	Expected m/z	rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		2
33	glucose-1-phosphate		259.0227	259.0225	6.61	0.69	16	16	15	5.3e+04	15	0.82		
32	AMP		346.0562 346.0561	346.0558 346.0558	15.32 11.57	4.02 0.27	16	14	39 59	1e+04 2.1e+04	7 1686	0.80 0.85		
31 30	AMP ribose-phosphate		229.0120	229.0119	7.55	1.55	16 16	16 5	35	2.1e+04 8.7e+03	1686 970	0.85		
29	ribose-phosphate		229.0120	229.0119	6.84	0.84	15	9	15	6.8e+03	830	0.85		
28	pantothenate		218.1034	218.1034	13.66	2.66	16	9	57	1.2e+04	11	0.82		
27	pantothenate		218.1034	218.1034	11.27	0.27	16	16	21	3.8e+06	34	0.84		
26 25	deoxyribose-phosphate D-glucarate		213.0170 209.0304	213.0170 209.0303	5.64 13.28	2.10 0.28	15 16	9 8	8 9	8.5e+03 6.1e+03	1160 500	0.85 0.84		0
25	D-giucarate tryptophan		209.0304 203.0825	203.0826	7.79	0.28	16	8 16	36	6.1e+03 1.5e+06	33	0.84		
							1.2	1.2						

Users can select All, Good, Bad or Selected peaks to export.

<u>File</u>	$\stackrel{Vidgets}{\equiv} \mathbf{k} [\mathbf{P} \times]$													
	Align Peaks Match Option:	5 Polly						Au	ito Detect(-1) Cha	arge 1 :AreaTop		[m/z] 218.10336304	+/- 5.000000	: ppm :
														C
đ 🗍			U 2				pa	ntothenate m/z: 2	218.1023-218.1045					*
S Sa	ample name filter	_4.8e+06									Deek Aree	Top		
	exampleProject	1.00000									3:88+88	1215 240M		հեր
odu	Ø 091215_240M													+++
5		_4.2e+06						~			2:28 88			\sim
-	○ 091215_120M ○ 091215_120i										3:48 88			
	Ø91215 45M ∠	_3.6e+06									2:88 88			20
		3.68+06									2:88106	91215_0		C°
	■ 091215_15M ■ 091215_15M ■ 091215_15i													۵.
	■ 091215 5M <	3e+06												
	■ 091215_5i <													<u></u> _
	Ø 091215_3M													-
	■ 091215 1M <	.2.4e+06						121						6
	■ 091215_1i <													
	Ø 091215_0M ∠ Ø 091215_0i ∠	1.8e+06												
	2 001210_01 C								A land					
		_1.2e+06												
		6e+05												
		00100												
		10 <mark>916</mark>	10.46		10.76		11.08	2	11.30	11.66		11.96	12.26	
Peak Table 1 🃭 🔗 🛠 🕴 🧏 🦂 🖓 🏗 🕎 🖨 🍺 👘														
# .	D	••• •0	O Export Select		rt	rt delta	#peaks	#good	Max Width	Max AreaTop	Max S/N	Max Quality		2
33	glucose-1-phosphate		25 Export All Gr	oups 0225	6.61	0.69	16	16	15	5.3e+04	15	0.82		
32	AMP		34 Export Good	0558	15.32	4.02	16	14	39	1e+04	7	0.80		
31 30	AMP ribose-phosphate		34 Export Bad 229.0120	0558 229.0119	11.57 7.55	0.27 1.55	16 16	16 5	59 35	2.1e+04 8.7e+03	1686 970	0.85 0.83		
29	ribose-phosphate		229.0120	229.0119	6.84	0.84	15	9	15	6.8e+03	830	0.85		
28	pantothenate		218.1034	218.1034	13.66	2.66	16	9	57	1.2e+04	11	0.82		_
27	pantothenate deoxyribose-phosphate		218.1034 213.0170	218.1034 213.0170	11.27 5.64	0.27 2.10	16 15	16 9	21 8	3.8e+06 8.5e+03	34 1160	0.84 0.85		
26 25	D-glucarate		209.0304	209.0303	5.64 13.28	0.28	15	8	8	6.1e+03	500	0.85		0
24	tryptophan		203.0825	203.0826	7.79	0.29	16	16	36	1.5e+06	33	0.84		
a a a b a bi	hanata													in l



The *Export Groups to CSV* option lets the users save the 'good'/'bad' labels along with the peak table. Users also have the option to filter out rows that have a certain label while exporting the table.



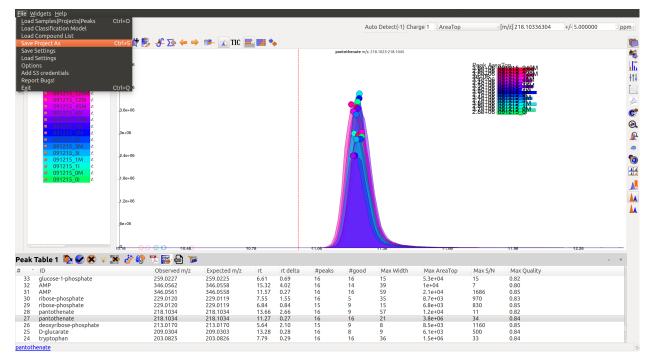
Generate PDF Report option

saves all EICs with their corresponding bar plots in a PDF file.



exports all EICs to a Json file.

Another option is to export the peak data in .mzroll format that can be directly loaded into El-MAVEN by clicking on the Load *Samples*|*Projects*|*Peaks* option in the File menu. For this, go to the File option in the menu bar, and click on 'Save Project'.



1.4.4 Validation

Labeled MS/MS Validation

MS/MS Support in EI-MAVEN

Tandem mass spectrometry or MS/MS is an important technique in analytical Mass spectrometry. Processing MS/MS data for a large batch of samples can be a time-consuming task. We have added some support for processing such data in El-MAVEN.

Data acquired using the following methods are currently supported in El-MAVEN: Multiple Reaction Monitoring, Parallel Reaction Monitoring (or Full-Scan MS2) and Data Dependant MS2.

Feature Additions

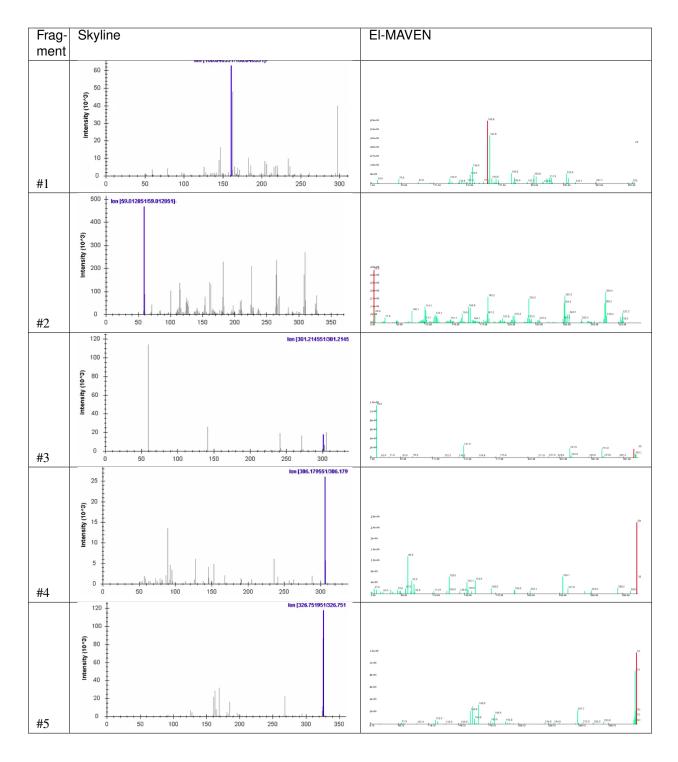
- Text search option using precursor and product m/z
- Targeted Peak Picking for MS/MS
- Manual annotation of SRM transition

Validation

In order to ensure correct reading and processing of MS2 sample files in El-MAVEN, MS spectra and chromatograms from El-MAVEN were validated using Skyline by ProteoWizard.

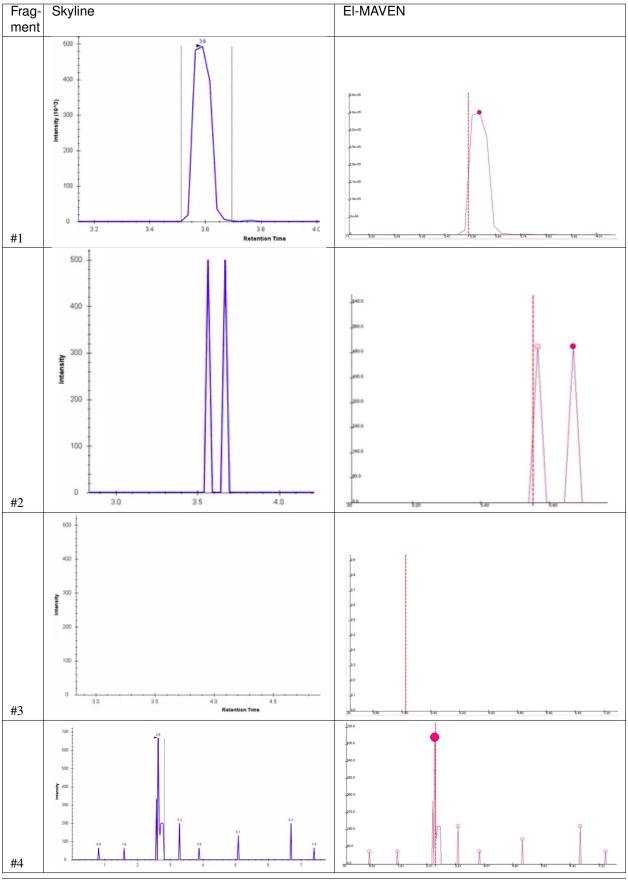
Mass Spectra Validation

Dataset: Full-scan MS2 sample file



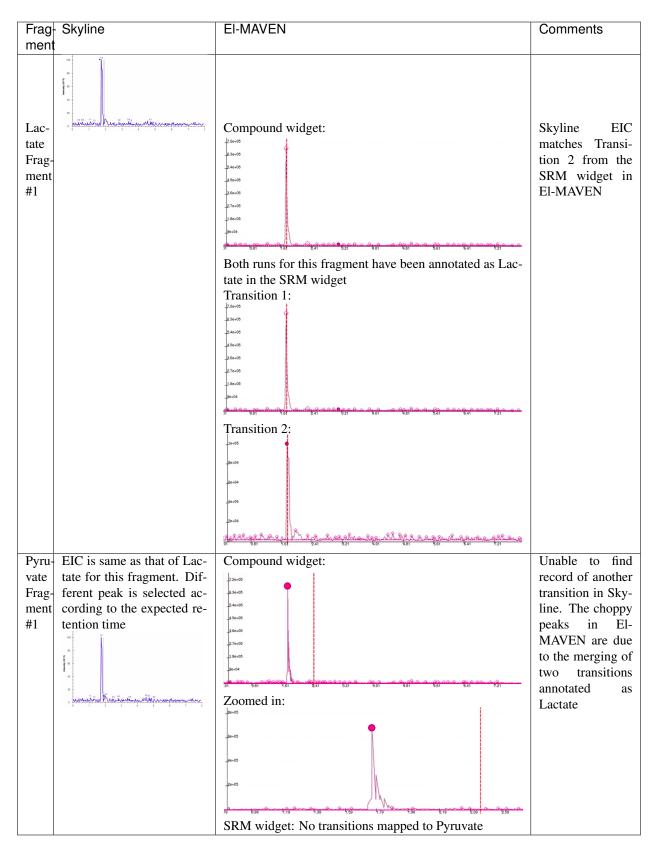
Chromatogram Validation

Dataset: Multiple Reaction Monitoring

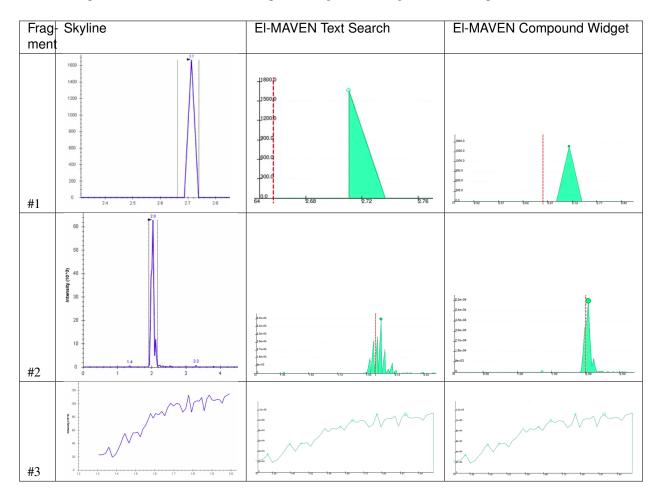


1.4. User Documentation

Multiple Transition cases: The same precursor/product pairs can be tracked in multiple runs, for example, in cases where two metabolites have common fragments like Pyruvate 89/89 and Lactate 89/89. El-MAVEN lists these runs separately in the SRM List widget. Following are some examples:

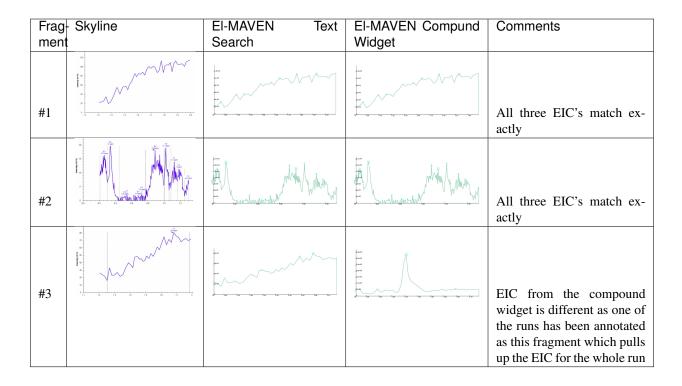


Text search/Compound widget mismatch: Using the text search feature for MRM data can result in choppy peaks at



times as compared to those from the SRM/compound widget. Following are some examples:

Dataset: Full-scan MS2



Issues Observed

- Text search and compound/SRM widget EIC mismatch for MRM data: EIC obtained from the compound widget and Skyline are the same. The data points on the plot are same for all three. The error in text search is due to different processing. This has been filed as issue number #487; Status: Unresolved
- Text search and compound widget EIC mismatch for PRM data: EIC from Text search and Skyline are the same. The error in compound widget happens due to linking of SRM ID to the compound widget. This pulls up the whole run instead of the particular fragment. This is a known issue #442; Status: Unresolved
- Merged peaks in case of multiple transitions: This is being handled as part of issue number #405; Status: In progress
- Mass spectra displays wrong product M/z value in some cases: The prodMz field in the title of spectra widget displays the base peak m/z value which may or may not be the same as the fragment m/z. This will be resolved as part of #396; Status: Unresolved
- Peak table displays NA in Expected m/z column for MRM data: This has been filed as issue number #493

Conclusion

Mass spectra validation against Skyline passed. This indicates correct parsing of MS/MS data in El-MAVEN. EIC validation against Skyline highlighted some known and unknown issues. We recommend using Text search for PRM data and Compound/SRM widget for MRM data processing until the reported issues have been fixed.

1.5 Contributors

• MAVEN team at Princeton University

- Eugene Melamud
- Victor Chubukov
- George Sabu
- Sahil
- Raghav Sehgal
- Shubhra Agrawal
- Raghuram Reddy
- Rishabh Gupta
- Pankaj Kumar
- Rahul Kumar
- Giridhari
- Kiran Varghese
- Naman Gupta
- Francis Lee
- Avijit Zutshi

1.6 Contributing

You are welcome to contribute. Please go through our contributing guidelines and code of conduct. These guidelines include directions for coding standards, filing issues and development guidelines.

1.7 References

To understand MAVEN's and El-MAVEN's workflows and features, please refer to the following literature on Maven:

- Metabolomic Analysis and Visualization Engine for LC-MS Data, Eugene Melamud, Livia Vastag, and Joshua D. Rabinowitz, Analytical Chemistry 2010 82 (23), 9818-9826
- LC-MS Data Processing with MAVEN: A Metabolomic Analysis and Visualization Engine, Clasquin, M. F., Melamud, E. and Rabinowitz, J. D. 2012, Current Protocols in Bioinformatics. 37:14.11.1-14.11.23.

1.8 Acknowledgement

El-MAVEN would not have been possible without the unwavering support, constant feedback and financial support from Agios Pharmaceuticals, Inc. El-MAVEN thanks the metabolomics community for its immense contribution in taking the tool forward and making it a great success.

1.9 Release History

1.9.1 Current Release

Version 0.4.1

Publication Date: August 4th, 2018

- Optimisations
 - Faster sample upload
 - Checking/unchecking a sample in the presence of a large peak table is faster (#723)
 - No lag in switching between two large peak tables
- UI Changes
 - Added a close button for Scatter Plot widget
 - Removed 'Load Peaks' from Peak tables
 - 'Gallery' widget is only available on the side bar panel
 - Removed 'Ratio' and 'p-value' columns from Peak tables (available in Scatter Plot Peak Table)
 - Removed 'Good' and 'Bad' marking options from EIC widget
 - Polly upload button is not clickable during the upload process (#805)
 - Single button for exporting EIC on EIC widget
 - Removed 'Scatter Plot' button from side bar panel
- Bugs
 - Isotopic peak filtering based on peak quality has been fixed (#772)
 - Crash in bookmark button has been fixed (#768)
 - Empty tables will no longer be uploaded to Polly (#777)
 - Only a single compound database file (.csv) will be uploaded to Polly

1.9.2 Past Releases

Version 0.4.0-beta.1

Publication Date: July 16th, 2018

- Fixed Isotopic detection in shift-drag integration (#781)
- Groups with zero peaks will not be bookmarked on shift-drag integration
- Fixed a bug in changing compound database from the drop-down
- Getting Started window will not open behind the main application (#775)

Version 0.4.0-beta

Publication Date: July 5th, 2018 Version 0.3.2 Publication Date: May 17th, 2018 Version 0.3.1 Publication Date: April 27th, 2018

- Bug Fixes:
 - Fixed compound name in CSV export (#635)
 - Fixed automated detection of high ranking groups
 - Fix installer issue on Mac (Installer not working due to netcdf addition)
- UI:
 - Change default tab for El-MAVEN Polly Integration dialog

Version 0.3.1

Publication Date: April 26th, 2018

Version 0.3.0

Publication Date: April 24th, 2018

Version 0.3.0

Publication Date: April 23rd, 2018

Version 0.3.0-beta

Publication Date: April 11th, 2018

- Features
 - Export Scatter plot groups into a peak table
 - Introduced Exception handling in sample upload process
 - Send logs to sentry in the event of a crash
 - Upload data to Polly through CLI and GUI
 - Download project settings from Polly
 - New Alignment algorithm: Obi-warp
 - Support for cdf files on Windows
 - Highlight detected compounds in the database
- Enhancement
 - Consistent sample ordering across widgets
 - Append Sample number to Sample Name for mzML files
- Refactor
 - Isotope Widget
 - Isotope detection
- Bug fixes
 - Fixed Isotopes expected m/z in mzroll
 - Blank samples will be visible in sample widget
 - Fix trailing delimiter in peak detailed format
 - Isotopic intensities are consistent between isotope widget and peaks table

Version 0.2.4 Publication Date: February 7th, 2018 Version 0.2.4 Publication Date: February 5th, 2018 Version 0.2.4 Beta Publication Date: January 24th, 2018 Version 0.2.3 Publication Date: January 23rd, 2018 Version 0.2.2 Publication Date: December 28th, 2017

1.10 Bugs and Feature Requests

Existing bugs and feature requests can be found on El-MAVEN's GitHub issue page. Please search the existing bugs and feature requests before you file one yourself.

1.11 Copyright and License

Code and documentation copyright 2017 Elucidata Inc. Code released under the GPL v2.0. Documentation is released under MIT license.