Supporting Information

Implementing Annotation Confidence Scoring in Untargeted Mass Spectrometry Workflows for Small Molecule Analysis

Daniel Krakko, Ralf Tautenhahn, Whitney L. Stutts January 9, 2026

Based on Schymanski et. al. https://pubs.acs.org/doi/full/10.1021/es5002105

Installing R, a free software environment for statistical computing and graphics

- Download (https://cloud.r-project.org/bin/windows/base/) and install **R**.
- Install the following libraries which are required to run this script: rjson, XML, RSQLite. You can do this by starting R (shown as R 4.5.1 in Windows Start menu) and **either**
  - use the menu in R (Packages -> Install Packages, find and select the packages rjson, XML and RSQLite in the very long list of packages that is displayed, click OK to install)
  - or by using the command below (recommended, faster). Copy and paste into R command prompt

```
install.packages(c("rjson", "XML", "RSQLite"))
```

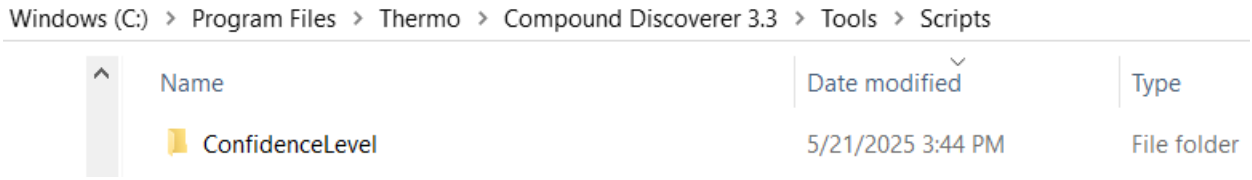
```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> install.packages(c("rjson", "XML", "RSQLite"))
```

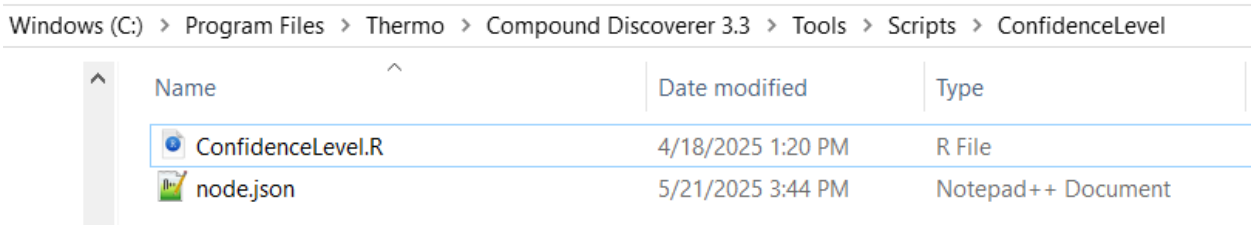
- Press Enter. R will ask you to select a CRAN mirror server. You can use the one that is preselected (0-Cloud) or any other server in the list and click OK. R will then install these 3 packages and potential dependencies. You can close the R application when the installation process is completed.

Installing the Scripting node

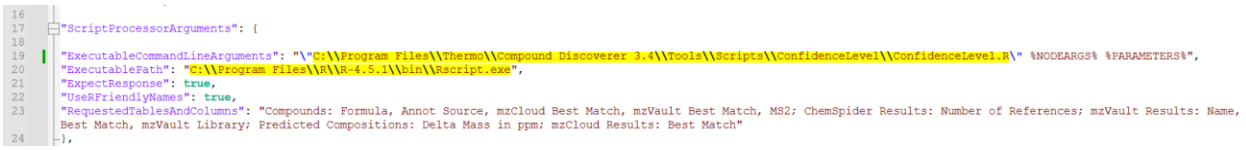
- Download the *ConfidenceLevel_CD3.x.zip* file that matches your CD version 3.x
- Note: some of the following steps will require administrator privileges.
- Use Windows Explorer to navigate to the following folder
  C:\Program Files\Thermo\Compound Discoverer 3.3\Tools\Scripts (for CD3.3)
  C:\Program Files\Thermo\Compound Discoverer 3.4\Tools\Scripts (for CD3.4)
  C:\Program Files\Thermo\Compound Discoverer 3.5\Tools\Scripts (for CD3.5)
- Copy the *ConfidenceLevel_CD3.x.zip* file you downloaded into this folder and extract it here. Do not create an additional subfolder.
- You should now see a folder “ConfidenceLevel”



- This folder contains two files, the actual R script as well as the *node.json* file, which describes certain properties of the scripting node, so that Compound Discoverer can install and use it.



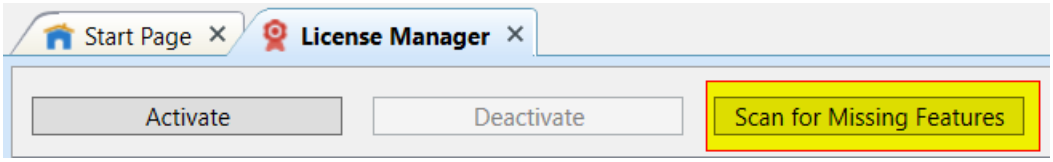
- If you are using the latest version of R (as of 8/27/2025 that is version 4.5.1) and you have copied the script into the folder exactly as described above, then no changes are needed in the node.json file. However, if you are using a different R version, if Compound Discoverer was installed into a different folder, or if the folders are named differently then you might have to update the node.json file. To do that, open the file in text editor like Windows Notepad or Notepad++.



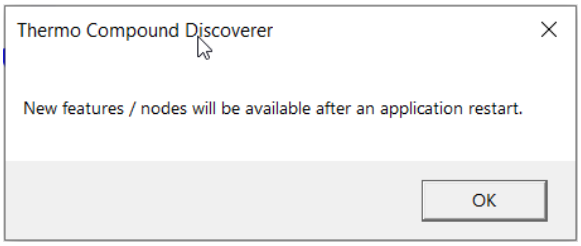
Line 19 contains the path to the R script. Edit path if necessary to match the location where you copied the script. Save and close the JSON file.

Line 20 contains the path to Rscript.exe. Edit path if necessary to match the R version that is installed on your PC (e.g. R-4.5.2 instead of R-4.5.1).

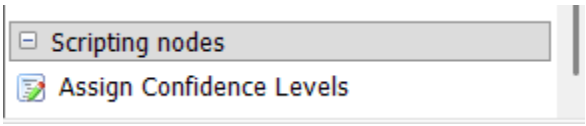
- To register the new node, start Compound Discoverer. Click on Help – License Manager in the main menu to open the License Manager.



- Click on **Scan For Missing Features**. After a few seconds, you should see the following information.

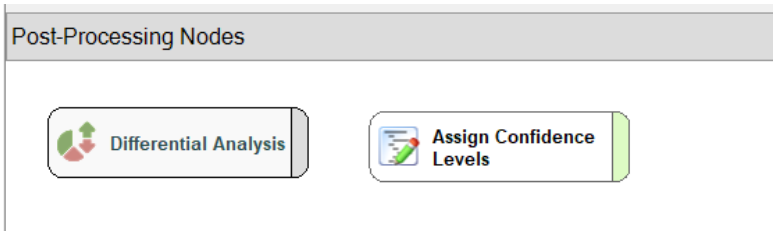


- Restart Compound Discoverer. You should now see the new node **Assign Confidence Levels** at the very bottom of your list of Workflow Nodes.

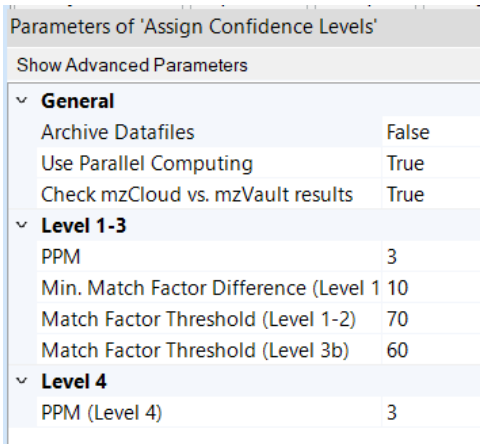


Using the new node in your workflow

The *Assign Confidence Levels* node is a post-processing node. Drag it into the Post-Processing Node area. Post-processing nodes are not connected to other nodes. The *Assign Confidence Levels* node requires that the Search mzCloud, Search mzVault, Search ChemSpider and Predict Compositions nodes are present in your workflow. It will also utilize information from other nodes, like BioCyc, mzLogic, Search MassLists etc. but the use of these nodes is optional. The *Assign Confidence Levels* node can be used with Reprocessing, that means the additional columns (see below) can also be added to already existing results.



Click on the *Assign Confidence Levels* node to see the parameters for this node.



Results

Name	Formula	Annot. Source	ConfLevel	ConfLevel_Flags	Redundant Annotation Flag	Annot. ΔMass [ppm]
DL-Arginine	C6 H14 N4 O2		2			-0.32
Stachydrine	C7 H13 N O2		1	L1 CpdName not matching mzVault w RT top Hit		-0.23
	C16 H9 N10 O14 P		4b			-0.45
Platelet-activating factor	C26 H54 N O7 P		3b			0.20
N-methylethanolamine phosphate	C3 H10 N O4 P		5			0.38
D-(+)-Pipecolinic acid	C6 H11 N O2		3a			-0.06
L-Propionylcarnitine	C10 H19 N O4		1	L1 CpdName not matching mzVault w RT top Hit		0.01
	C23 H15 N6 O15 P S		4b			1.81
L-Phenylalanine	C9 H11 N O2		1			-0.05
1-[(9Z)-hexadecenoyl]-sn-glycero-3-phosphocholine	C24 H48 N O7 P		3b			0.46
2,2-Dimethylpiperazine-1,4-diol	C6 H14 N2 O2		3a			-0.27
Indoline	C8 H9 N		3b			0.17
(4S)-4-[(2E)-2-Octenoyloxy]-4-(trimethylammonio)buta	C15 H27 N O4		3b			0.25
	C34 H82 N9 O6 P3		4a			2.60
Norcocaine	C16 H19 N O4		2			0.34
Isobutyryl-L-carnitine	C11 H21 N O4		3a	L1 CpdName not matching mzVault w RT top Hit		0.14
	C22 H51 N5 O6 P2		4b			2.06
	C6 H11 N2 O P S		4a			0.27
Hypaphorine	C14 H18 N2 O2		2			0.77
Prolylleucine	C11 H20 N2 O3		2		L2 duplicate-identical evidence	0.22

The additional columns *ConfLevel*, *ConfLevel_Flags*, and *Redundant Annotation Flag* are shown in the Compounds table.

The script also generates an additional table, *ConfidenceLevelSummary*, which provides a summary of the count of compounds observed at each confidence level.