# Large-scale analysis of non-targeted LC-MS metabolomics data with OpenMS in the Compound Discoverer™ platform

**Fabian Aicheler**[1], Timo Sachsenberg[1], Erhan Kenar[1], Sebastian Kusch[2], Hans Grensemann[2], Oliver Kohlbacher[1]
[1]Center for Bioinformatics, Tübingen, Germany; [2]Thermo Fisher Scientific GmbH, Germany

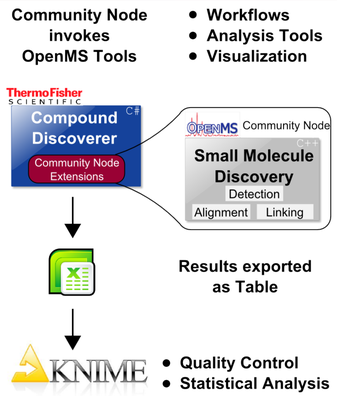
## Overview

**Purpose**: Integration of a pipeline for feature identification and quantification into Thermo Fisher Compound Discoverer™ (CD).
**Implementation**: Performance of a feature detection algorithm was demonstrated on dilution series. An OpenMS [1] pipeline was constructed around this algorithm and wrapped into a CD community node. Pipeline output was incorporated into the CD reporting format.
**Results**: We announce the first integration of an automated workflow for metabolite quantification into the novel CD platform, providing a community extension that enables the differential analysis of multiple LC-MS runs.

## Introduction

- Label-free quantification of small molecules using LC-MS has become a standard analytical technology.
- Complex LC-MS datasets require automated processing such as mass trace detection and assembly of isotopic traces to features, followed by quantification and identification of compounds.
- Recently, Kenar et al. [2] presented a sensitive feature detection algorithm which results in reproducible metabolite quantification for small molecules.
- CD is a new mass spectrometry analysis platform scheduled for a release later this year. Analogous to Proteome Discoverer, which is tailored to proteins, CD is adapted for small molecule analysis. CD has been designed to allow integration of external tools and algorithms as so-called community nodes.
- Our aim was the integration of a metabolic feature identification and quantification pipeline into CD. Besides creating the necessary interfaces, consistent presentation of CD results and their export for straightforward downstream analysis outside of CD were declared goals.

## Implementation




Sources: tinyurl.com/lm85xsu, tinyurl.com/ldtvazo.

**Figure 1:** The OpenMS pipeline is implemented as community node in CD. Result export as tables allows downstream analysis outside CD, for example in Knime.

The open-source software library OpenMS allows for rapid development of mass spectrometry algorithms and tools. It includes methods for retention time alignment and feature linking [3]. We expanded this toolset with a novel algorithm for feature detection and non-targeted quantification of small molecule LC-MS data. Our OpenMS metabolite quantification workflow was encapsulated in a single CD node. Evaluations of the method by Kenar et al. included human plasma samples with spiked-in metabolites.

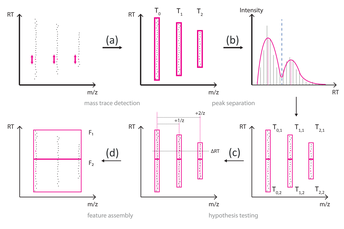



**Figure 2:** Methodology overview of the feature detection algorithm.

## Results

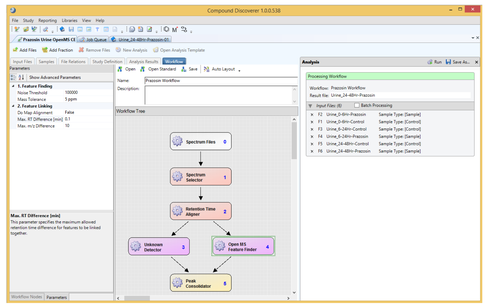


**Figure 3:** OpenMS community node in a CD example workflow.

OpenMS results are incorporated into the data processing and visualization capabilities of CD, offering tightly integrated presentation and downstream analysis in the CD software. Restriction to Thermo Fisher instruments allowed optimized parameter choices for the OpenMS algorithms.

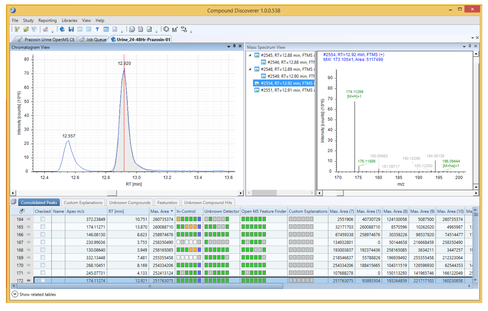


**Figure 4:** Result view of CD with integrated OpenMS results.

In the evaluation of the feature detection algorithm, correlations above **0.98** between feature intensities and corresponding compound concentrations were reported.

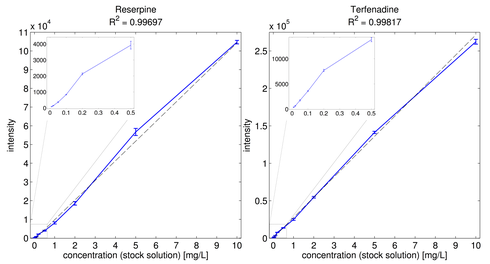



**Figure 5:** Correlations for chosen metabolites in dilution experiments.

To assess the quality of our small molecule detection pipeline, we investigated reproducibility in terms of feature recurrence over multiple measurements. Our integrated feature detection algorithm was compared with XCMS/CAMERA in a dilution series (33 MS runs). A time series of Prazosin metabolism in rats was used to compare our method with the feature detection algorithm provided by CD (6 MS runs).

| Found in # samples | OpenMS | XCMS Camera |
|---|---|---|
| 1-5 | 5590 | 1837 |
| 6-9 | 591 | 242 |
| 10-13 | 258 | 115 |
| 14-17 | 183 | 78 |
| 18-21 | 124 | 82 |
| 22-25 | 124 | 52 |
| 26-29 | 128 | 52 |
| 30-33 | 744 | 341 |

| Found in # samples | OpenMS | Component Elucidator |
|---|---|---|
| 1 | 5369 | 5777 |
| 2 | 2480 | 2684 |
| 3 | 1719 | 1434 |
| 4 | 2288 | 1543 |
| 5 | 1491 | 903 |
| 6 | 1446 | 1173 |

**Table 1: Left:** Reproducible features for OpenMS and XCMS/CAMERA (dilution series, 33 MS runs). **Right:** Reproducible features for OpenMS and Component Elucidator (Prazosin time series, 6 MS runs).

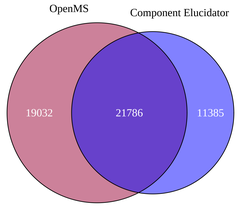



**Figure 6:** Overlap of detected features measured in a time series of Prazosin metabolism.

CD results can be exported to tabular file formats, allowing downstream analysis outside of CD. Here we use the KoNstanz Information MinEr (KNIME) [4]. KNIME supports a multitude of processing modules for cheminformatics, machine learning and statistics. A downstream analysis workflow in KNIME (See Figure 7) allows elaborate analysis of CD results.

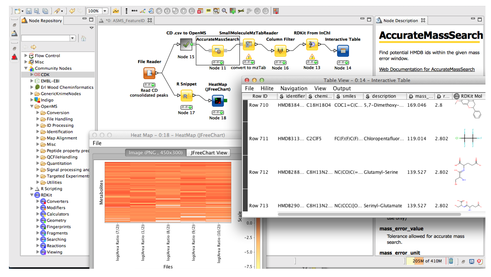



**Figure 7:** Example analysis in KNIME.

## Conclusion

- We successfully integrated a robust, sensitive feature quantification method into CD, enabling joint analysis of multiple runs.
- Reduction of parameters and integration into CD significantly improved accessibility of this state of the art metabolite quantification workflow.
- Source code (C#) of our community node will be freely available under an open-source license parallel to the release of CD.

## References

[1] Sturm et al. OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinformatics*, 9:163, 2008.

[2] Kenar et al. Automated label-free quantification of metabolites from liquid chromatography-mass spectrometry data. *Mol. Cell. Proteomics*, 13(1):348–59, 2014.

[3] Weisser et al. An automated pipeline for high-throughput label-free quantitative proteomics. *J. Proteome Res.*, 12(4):1628–1644, 2013.

[4] Berthold et al. KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Springer, 2007.