

HowDirty: An R package to evaluate molecular contaminants in LC-MS experiments

David Gomez-Zepeda¹, Thomas Michna², Tanja Ziesmann², Ute Distler², and Stefan Tenzer²

¹Helmholtz-Institute for Translational Oncology Mainz (HI-TRON), Germany

²University Medical Center of the Johannes Gutenberg University Mainz

July 26, 2023

Abstract

Contaminants derived from consumables, reagents, and sample handling often negatively affect LC-MS data acquisition. In proteomics experiments, they can markedly reduce identification performance, reproducibility, and quantitative robustness. Here, we introduce a data analysis workflow combining MS1 feature extraction in Skyline with HowDirty, an R-markdown-based tool, that automatically generates an interactive report on the molecular contaminant level in LC-MS data sets. To facilitate the interpretation of the results, the HTML report is self-contained and self-explanatory, including plots that can be easily interpreted. The R package HowDirty is available from <https://github.com/DavidGZ1/HowDirty>. To demonstrate a showcase scenario for the application of HowDirty, we assessed the impact of ultrafiltration units from different providers on sample purity after filter-assisted sample preparation (FASP) digestion. This allowed us to select the filter units with the lowest contamination risk. Notably, the filter units with the lowest contaminant levels showed higher reproducibility regarding the number of peptides and proteins identified. Overall, HowDirty enables the efficient evaluation of sample quality covering a wide range of common contaminant groups that typically impair LC-MS analyses, facilitating taking corrective or preventive actions to minimize instrument downtime.

Hosted file

HowDirty_manuscript_proteomics_20230724.docx available at <https://authorea.com/users/643346/articles/656759-howdirty-an-r-package-to-evaluate-molecular-contaminants-in-lc-ms-experiments>