

## Contents

|   |   |
|---|---|
| 1. Introduction.....                              | 1 |
| 2. Requirements .....                             | 1 |
| 3. Procedure .....                                | 2 |
| 3.1. MS1 feature extraction in Skyline.....       | 2 |
| 3.2. Contamination assessment using HowDirty..... | 5 |
| 4. References .....                               | 7 |

---

## 1. Introduction

|                   |   |
|-------------------|---|
| <u>Objective:</u> | Assess the level of contamination of LC-MS results using the HowDirty workflow.   |
| <u>Input</u>      | LC-MS raw files + Target List from publication (see instructions where to find it)  |
| <u>Output:</u>    | HTML report with plots evaluating the normalized abundance of contaminants in the samples. Excel file compiling the results and summary statistics.   |
| <u>Summary:</u>   | The presence of contaminants (e.g., PEG) and detergents (e.g., CHAPS, SDS) in samples analyzed by LC-MS can be severely detrimental to identifying peptides/proteins or other molecules. Skyline is used to extra MS1 features of many known contaminant masses from raw files (e.g., .raw and .d). The results are exported to a .csv file, then processed in R using HowDirty to generate an HTML interactive report that evaluates sample contamination risks. |

Please cite

---

## 2. Requirements

- Raw LC-MS results to be evaluated
- Skyline version > 4 [1,2].  
To install it, you can register online and download the latest version here:  
<https://skyline.ms/project/home/software/Skyline/begin.view>.
- Skyline HowDirty template, including the Skyline molecular contaminant transition list [3] and reports configuration: <https://github.com/DavidGZ1/HowDirty/tree/main/tutorial>
- Alternatively, you can set up Skyline yourself (further instructions in the [Skyline tutorials](#))
  - Download the original template with the molecular contaminant transition list [3] from [Panorama](#) and load it into Skyline: **File / Open**
  - Create the PeakAreas\_Contaminants report: containing the columns:
  - **Settings / Document Settings / Reports / Add**, then enter the name "PeakAreas\_Contaminants" and select the columns (Tip: click on the binoculars symbol to access the search):  
"Proteins", "Peptides", "Replicate Name", "Peptide Retention Time", "Total Area MS1",  
"Total.Ion.Current.Area"
  - Enable the report form by ticking the box next to its name, then click **OK**
- R software for data analysis and the R packages Rmarkdown, knitr, and HowDirty  
[HowDirty](#) can be installed by running the following code in R using a freshly open console window (e.g., in R Studio):

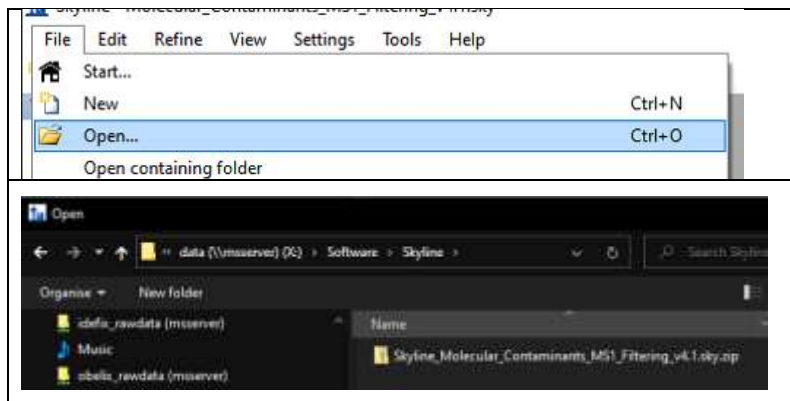
```
if(!require(devtools)) install.packages("devtools")
devtools::install_github("kassambara/ggpubr")
devtools::install_github("DavidGZ1/HowDirty", force = TRUE)
```

### 3. Procedure

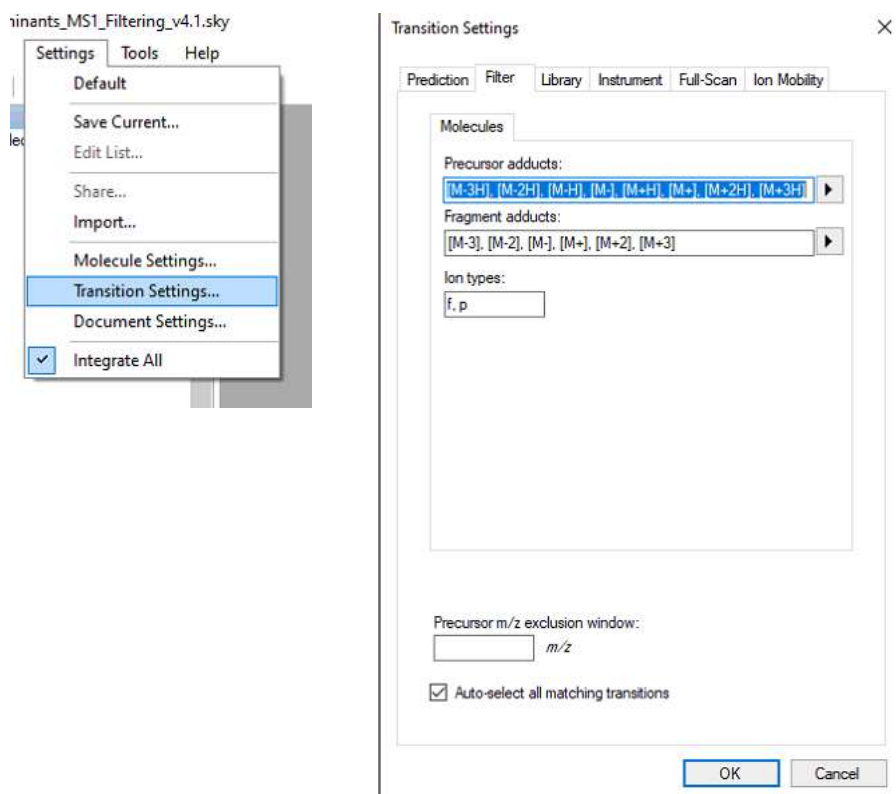
#### 3.1. MS1 feature extraction in Skyline

1. Open Skyline
2. Open the HowDirty template: **File / Open** navigate to the file location and open the .sky file

A Target List of commonly known contaminant masses will appear in Skyline.



3. Save the Skyline's file in the desired location
4. Navigate to **Settings / Transition Settings / Filter** and **/ Full-Scan** and set the parameters accordingly to the experiment. An example for Exploris 480 data is shown below.



Filter settings must include "p" (precursor) in Ion Types.



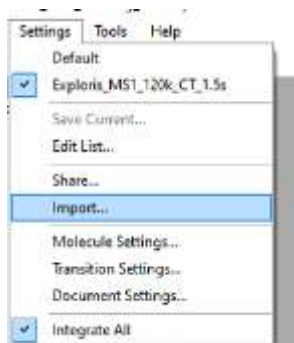
Instrument, Full-Scan, and Ion Mobility depend on the instrument and method used

Resolving power may depend on method-specific parameters. For instance, Exploris 480 can be acquired with an MS1 resolving power of 60000 at 200 m/z. and MS2 15,000 at 200 m/z.

For TIMS-ToF data, choose “TOF” as the precursor mass analyzer and resolving power 38.000 in both corresponding fields.

**Important:** in retention time filtering, select “include all matching scans”

5. Optional: Save those settings so that they can be reused by importing.

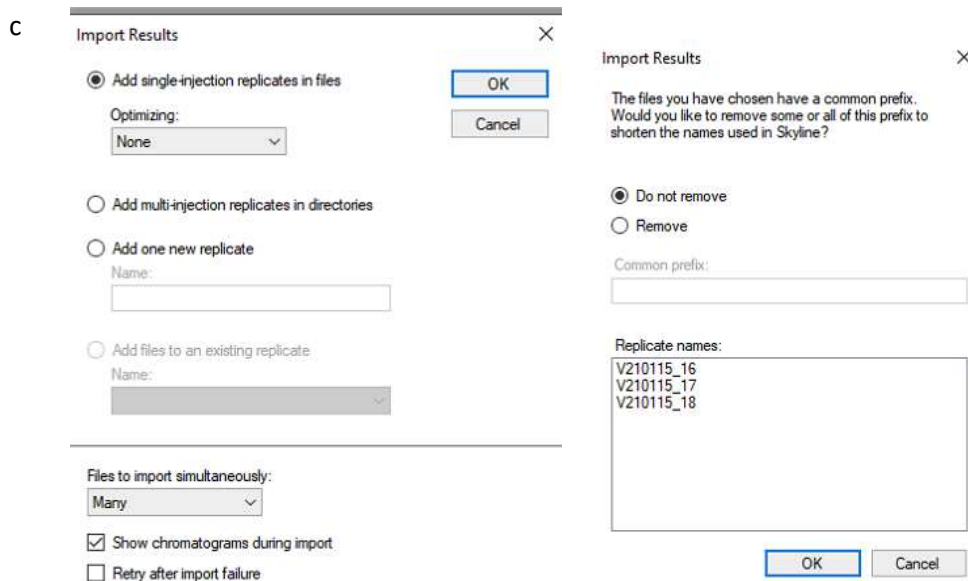
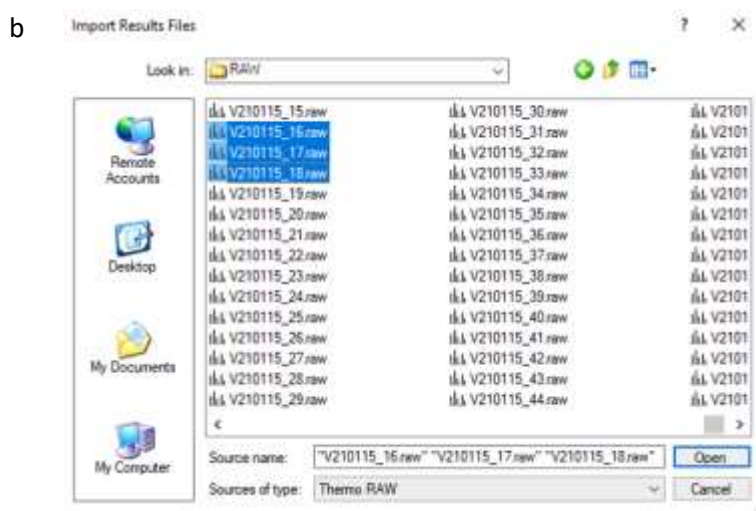
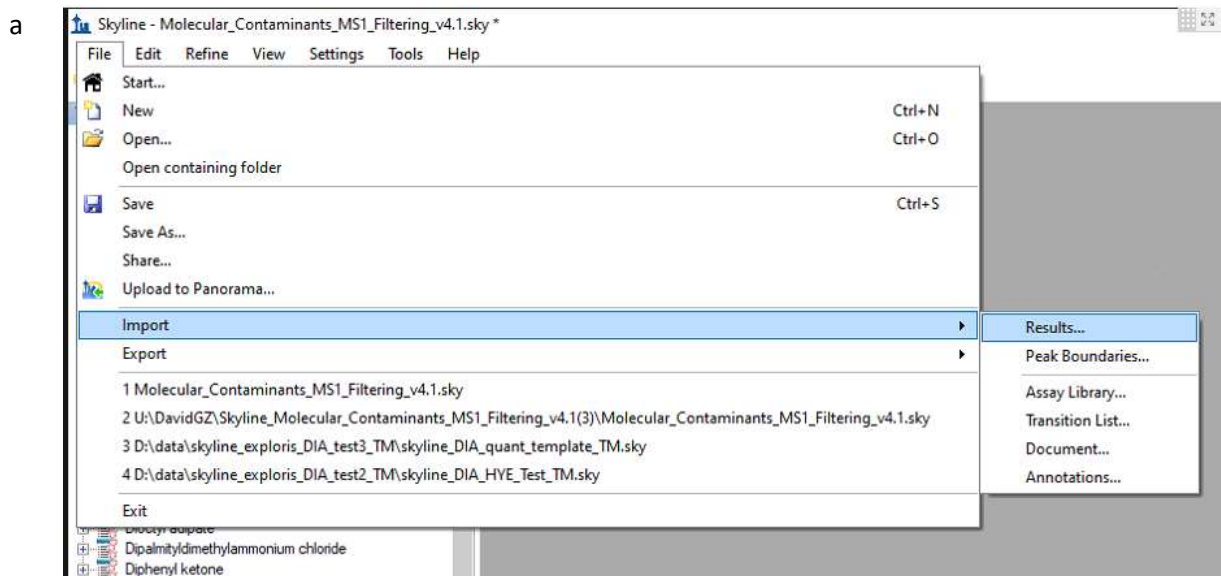


Use *Settings/Share* to export.

Use *Settings/Import...* to import.

6. Import the raw data (screenshots on next page)

- Select **File / Import / Results...**
- Select the raw files and click **Open**
- Select “add single-injection replicates in files” and click **OK**, then add file name pattern to remove if you want to do it and select **Remove**, or select **Do not remove / OK** if not

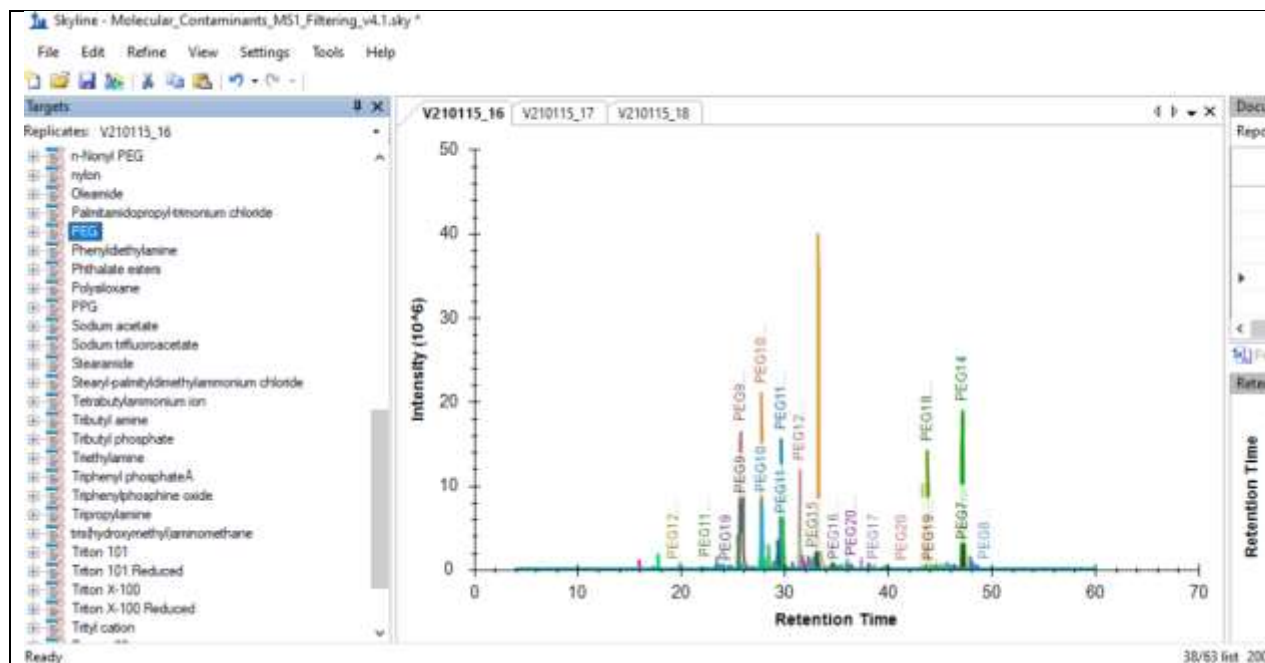


- After the processing of raw files is finished with processing, you can browse through the list of contaminants and see the respective extracted ion chromatogram (EIC/XIC).

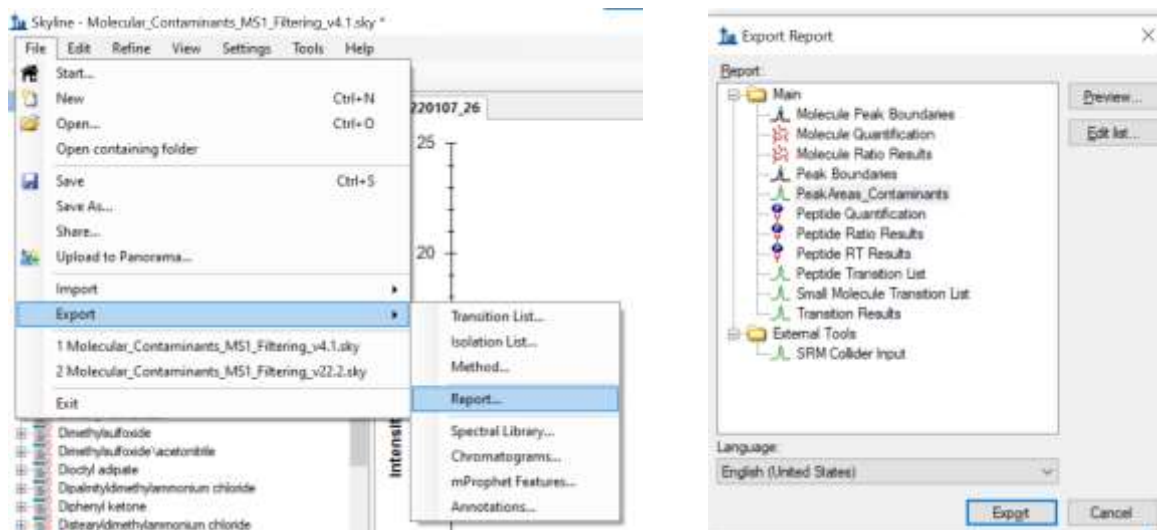
*You can do a first assessment at this level; for instance:*

*Does PEG elute in regular patterns in the chromatogram?*

*Does CHAPS elute with late retention time in the chromatogram?*



- Export the report by selecting **File / Export / Report...**, select the report PeakAreas\_Contaminants, navigate to the desired location and export the report as .csv (Note: use Language = English)



### 3.2. Contamination assessment using HowDirty

- Open R in your IDE of preference (e.g., RStudio) and set the working directory to the desired folder (e.g., where the PeakAreas\_Contaminants.csv is stored).

```
setwd("C:/Users/Name/ExampleHowDirty")
```

Alternatively, from RStudio, you can create a project: **File / New Project / Existing Directory / Browse / (select location) / Create Project**

- Load the HowDirty package

```
library(HowDirty)
```

11. Create the sample annotation or experiment design file using the code below. If a file\_report\_skyline is provided (PeakAreas\_Contaminants.csv), the ReplicateNameSkyline and Sample columns are populated with the unique(input\$Replicate.Name). If not, only the column headers are provided.

```
get_annotation_template(file_report_skyline = "PeakAreas_Contaminants.csv")
```

12. Open the samples\_annotation file in Excel (or R) and modify it if necessary.

ReplicateNameSkyline: must contain the Replicate.Name entries from the Skyline output

*Do not modify it if the HowDirty template is generated as described in point 13.*

Condition: Condition or group of each sample. Conditions are used for grouped statistics and plots.

Sample: sample name or code can be different from ReplicateNameSkyline

DilutionFactor: If the sample was diluted, this factor can be used to multiply the abundance results. Use carefully since abundance may not scale in the same proportions for all the contaminants.

13. Create a HowDirty template by using the code below; this will automatically save the template with the name entered (e.g., "example.Rmd")

```
HowDirty::get_template(file = "example")
```

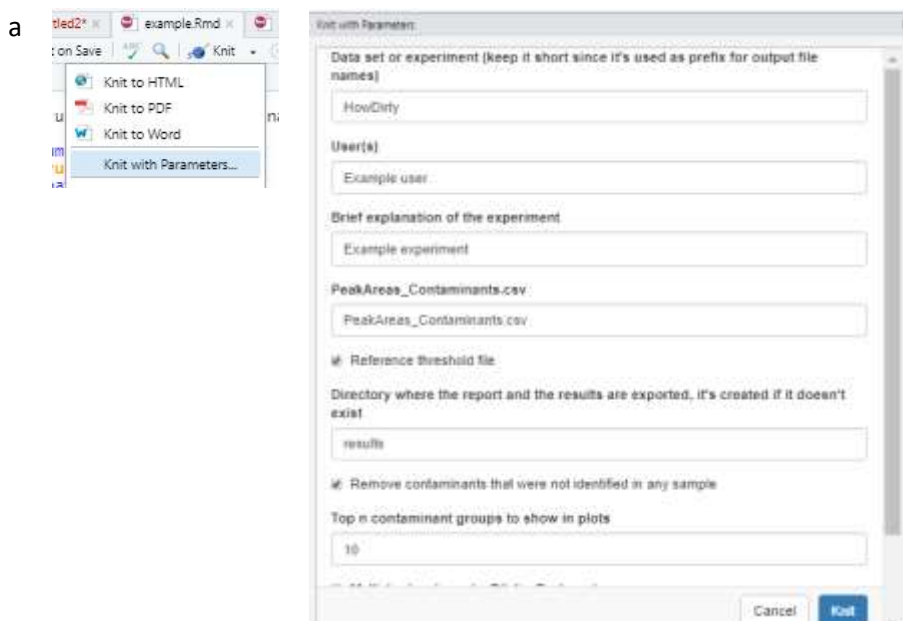
*Alternatively, open a HowDirty.Rmd file from previous analysis and save it with another name*

14. Open the .Rmd file

15. Fill the parameters in the header and "knit" (compile) the report. There are two options for this (see screenshots below).

*This may take some minutes, depending on the computer used and the number of samples*

- Click on "Knit with parameters"
- Fill out manually, then click "Knit"  
The parameters are in the header of the document after "params:"  
The values are entered **after "value:"**





b

```
title: evaluation of LC-MS contamination risk (HowDirty)
output:
  html_document:
    toc: true
    toc_float: false
    theme: united
    number_sections: true
params:
  datasets:
    label: "Data set or experiment (keep it short since it's used as prefix for output file names)"
    value: HowDirty
    input: text
  usernames:
    label: "User(s)"
    value: Example user
    input: text
  notes:
    label: "Brief explanation of the experiment"
    value: Example experiment
    input: text
  peakAreasContaminantsFile:
    label: "PeakAreas_Contaminants.csv" # PeakAreas_Contaminants report (directory/name) resulting from skyline
    value: PeakAreas_Contaminants.csv
  annotationFile:
    label: "Sample annotation file" # Samples Annotation directory/name
    value:
  refThresholdFile:
    label: "Reference threshold file" # file directory/name or FALSE
    value: NULL
  outputDirectory:
    label: "Directory where the report and the results are exported, it's created if it doesn't exist"
    value: results
  removeMissingContaminants:
    label: "Remove contaminants that were not identified in any sample"
    value: TRUE
  nTopContaminantGroups:
    label: "Top n contaminant groups to show in plots"
    value: 10
    input: numeric
  multiplyDilutionFactor:
    label: "Multiply abundance by dilutionFactor column" # use carefully, it multiplies the abundance by the dilutionFactor
    value: FALSE
  plotsInteractive:
    label: "Export interactive plots" #TRUE = interactive in html, FALSE = static
    value: TRUE
```

16. When the report is finished, it will open in R. You can navigate through it in that window or open it in any browser.

## 4. References

- [1] L.K. Pino, B.C. Searle, J.G. Bollinger, B. Nunn, B. MacLean, M.J. MacCoss, The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics, *Mass Spectrom. Rev.* 39 (2020) 229–244. <https://doi.org/10.1002/mas.21540>.
- [2] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, *Bioinformatics.* 26 (2010) 966–968. <https://doi.org/10.1093/bioinformatics/btq054>.
- [3] M.J. Rardin, Rapid Assessment of Contaminants and Interferences in Mass Spectrometry Data Using Skyline, *J. Am. Soc. Mass Spectrom.* 29 (2018) 1327–1330. <https://doi.org/10.1007/s13361-018-1940-z>.