# Getting streamlined elaboration of feature tables with separated quality controls, advanced statistics such as linear model with mixed effects, and more: presenting the "GetFeatistics" R-package

Gianfranco Frigerio

Any questions? Write me!!! ☺

**frigerio.gianfranco@hsr.it**

**Gianfranco Frigerio**

**@gianfrancofrig1**

Gianfranco Frigerio[1,2,3]

1 Center for Omics Sciences (COSR), IRCCS San Raffaele Scientific Institute, Milan, Italy

2 Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Belvaux, Luxembourg

3 Department of Clinical Sciences and Community Health, University of Milan, and Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

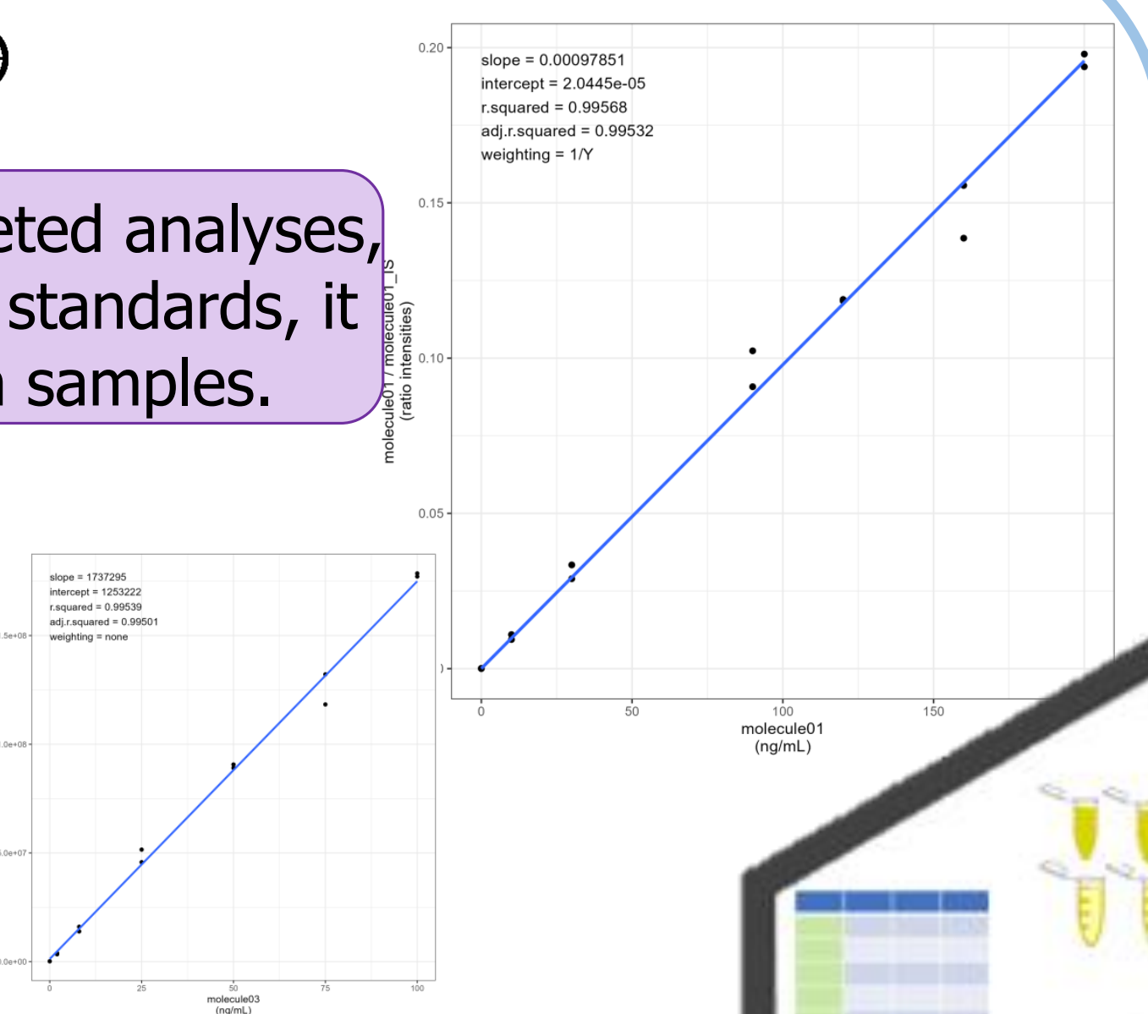
## Why this package?

Within the metabolomics community, tools such as MS-DIAL and R-based algorithms as XCMS are widely implemented to convert raw data to feature tables. As demonstrated in a previous work, while conducting non-targeted metabolomics studies with defined groups of subjects assessing variations of molecules present at low concentrations (such as metabolites of exogenous compounds) the use of separated pooled quality control (QC) samples can be a promising strategy to both improve the quality of the dataset and preserving potential features deriving from low concentrated molecules. Moreover, in observational epidemiologic studies, controlling for confounding factors is crucial to assess the association of metabolite variations with the biological question of interest; also, for longitudinal studies, linear models with mixed effects are a great statistical approach.

## Workflow

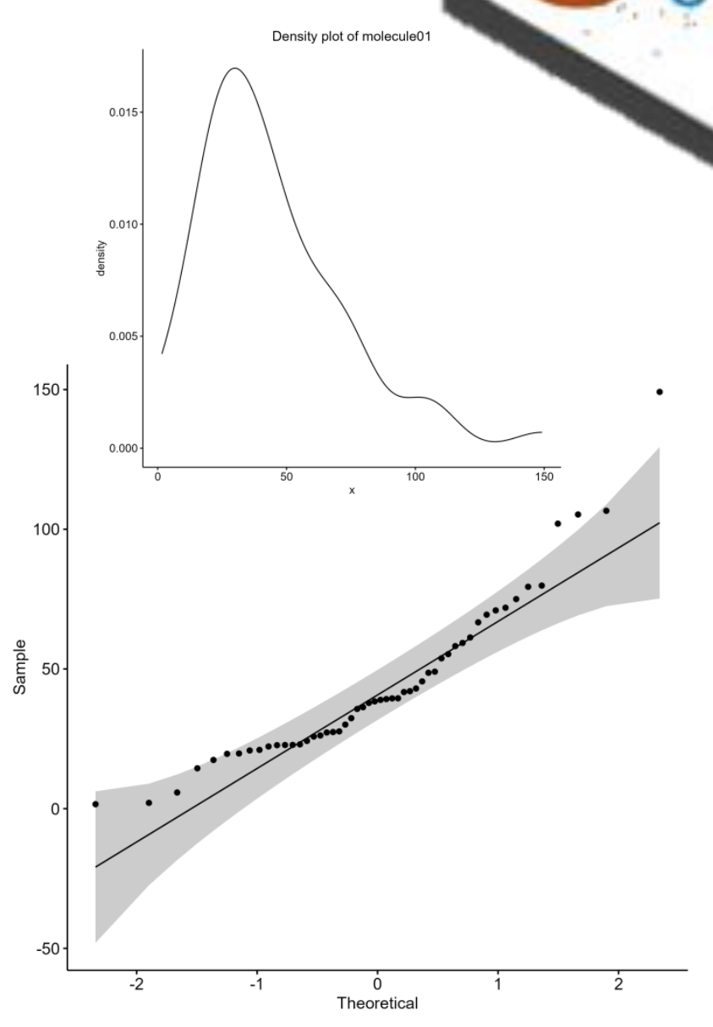
### Targeted analyses

| | |
|---|---|
| get_targeted_elaboration | Given a tables containing data from targeted analyses, including a calibration curve and internal standards, it calculates the concentrations in unknown samples. |
| plot_calibration_curves | It creates plots of the calibration curves from a targeted elaboration. |



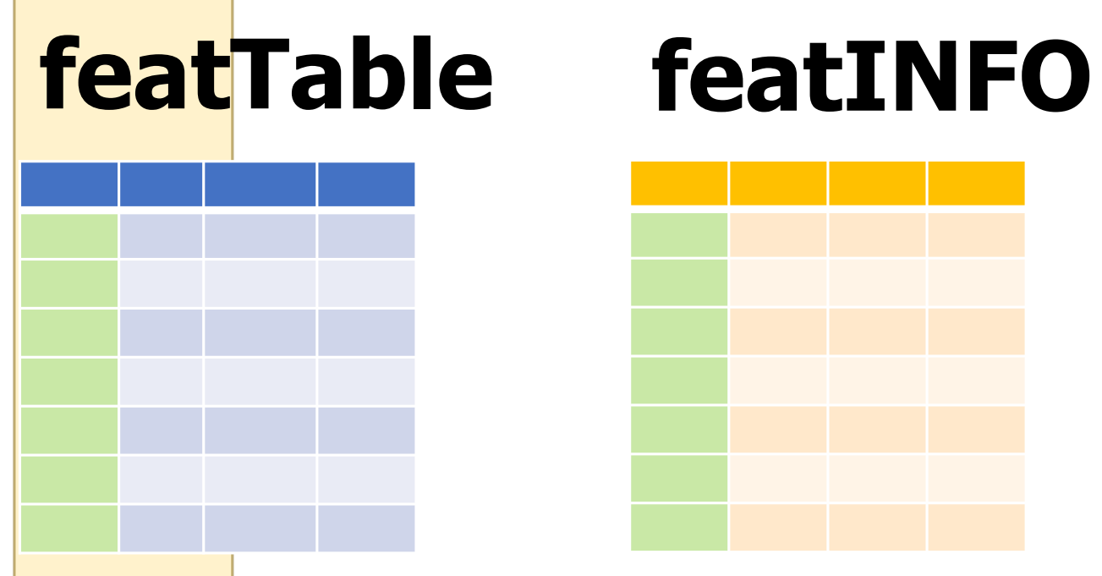
### Statistical analyses

| | |
|---|---|
| test_normality_density_plot<br>test_normality_q_q_plot<br>test_normality_saphiro_table | Functions to test normality, by producing density plot, q-q-plot and performing Shapiro test |
| transf_data | To replace missing values, log-transform, and/or scale the data. auto_scale, pareto_scale, mean_scale, range_scale |
| gentab_descr | generate a table with descriptive statistics for each of those variables |
| gentab_P.t.test | It performs t-test to each desired variable and creates a new table with the *p*-values |
| gentab_P.1wayANOVA_posthocTurkeyHSD | It performs one-way ANOVA, with also TurkeyHSD posthoc tests for between-groups comparison, to each desired variable and creates a new table with the *p*-values |
| gentab_P.2wayANOVA_posthocTurkeyHSD | It performs 2-way ANOVA, with also TurkeyHSD posthoc tests for between groups comparison, to each desired variable and creates a new table with the *p*-values<br>• Interactions test can also be performed |
| fix_names | It removes special characters from a character vector |
| addINFO_to_table | Given a table containing data related to features (for example the output of a statistical analyses) and the related featINFO table, it combines everything in a single table. |



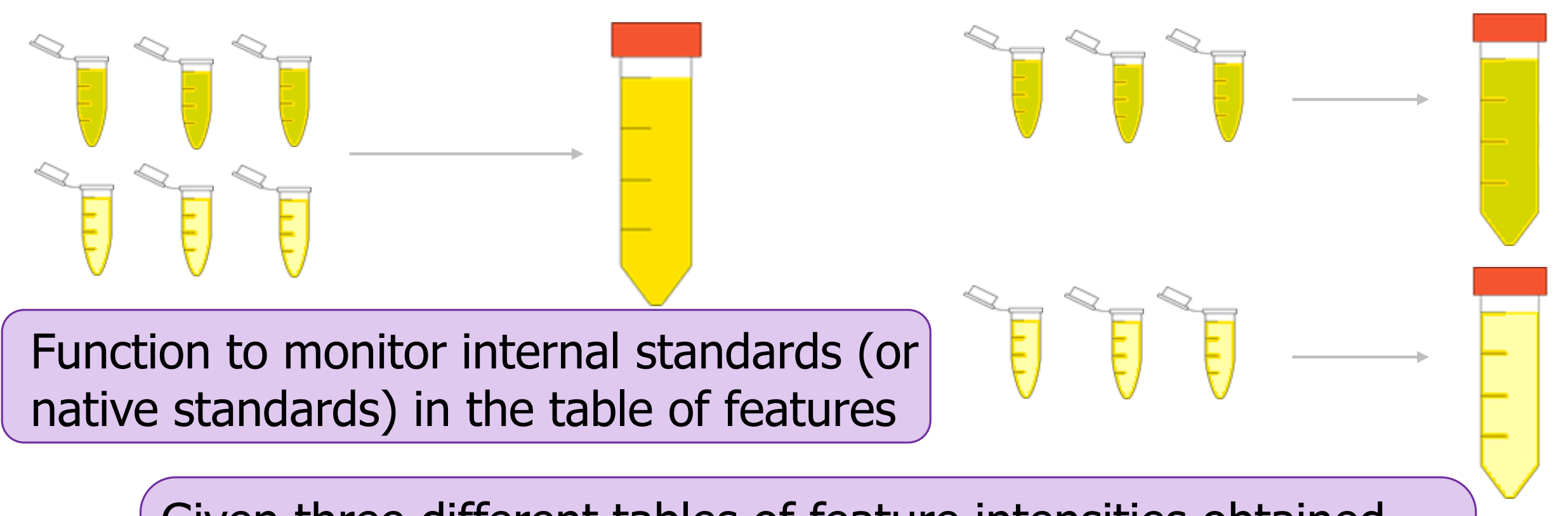
### Non-targeted analyses

| | |
|---|---|
| get_feat_info_from_MSDial<br>get_feat_info_from_patRoon<br>get_feat_table_from_MSDial<br>get_feat_table_from_patRoon | From MS-DIAL output or from patRoon output, it creates the "featTable" and the "featINFO" |



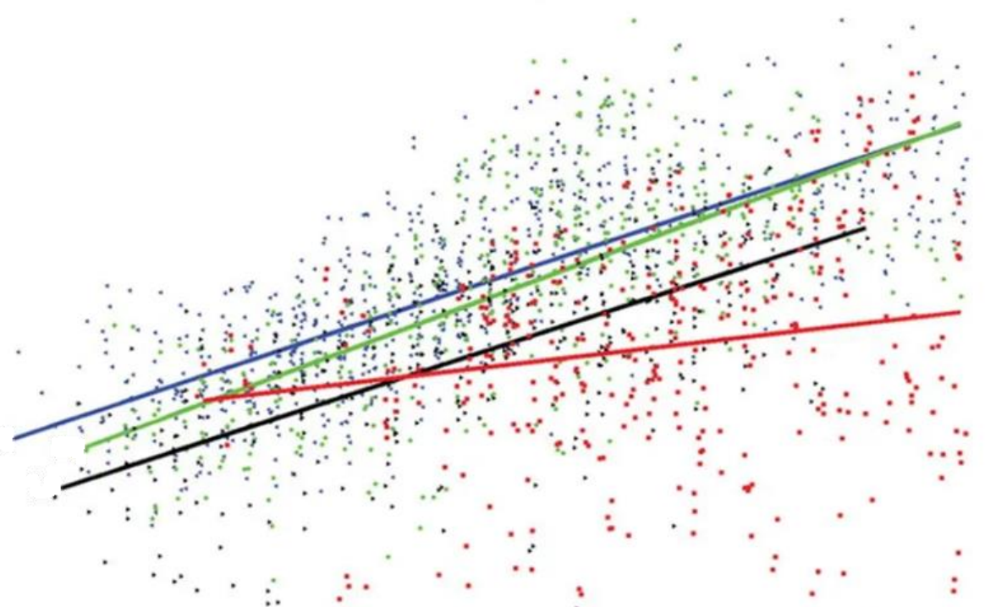
- **featTable:** table with feature intensities
- **featINFO:** table with rt, m/z and potential annotations

| | |
|---|---|
| get_AnnoLevels_MSDial<br>get_AnnoLevels_patRoon | Given pre-defined cut-off values, it assigns annotation levels to each feature to featINFO |
| QCs_process | It filters the featTable intensities considering some defined cut-off in the quality control samples. Including separated pooled quality control per sample group |
| *checkmolecules_in_feat_table* | Function to monitor internal standards (or native standards) in the table of features |
| GetCombined_featTable_MSDial | Given three different tables of feature intensities obtained from different batches, and the related featINFO tables, it combines the feature intensities in a single table. |
| transpose_feat_table | Given a feature table, it transposes it, so it is more suitable for following elaborations and statistical analyses (samples in rows and features in columns). |

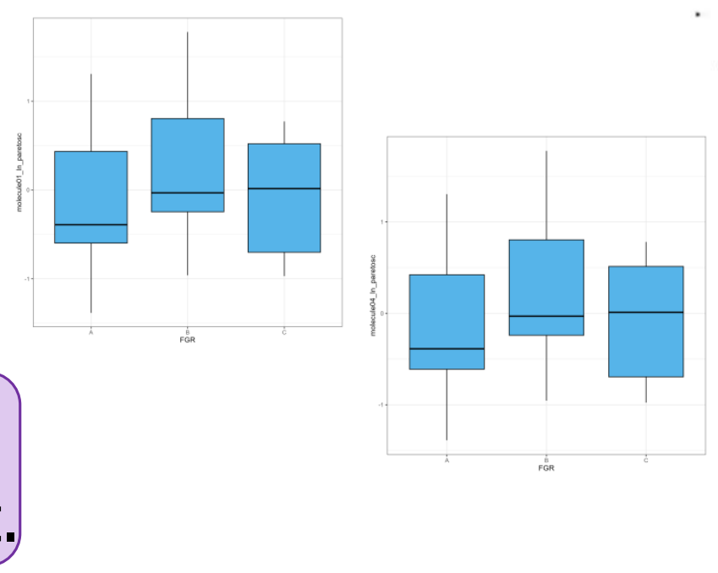


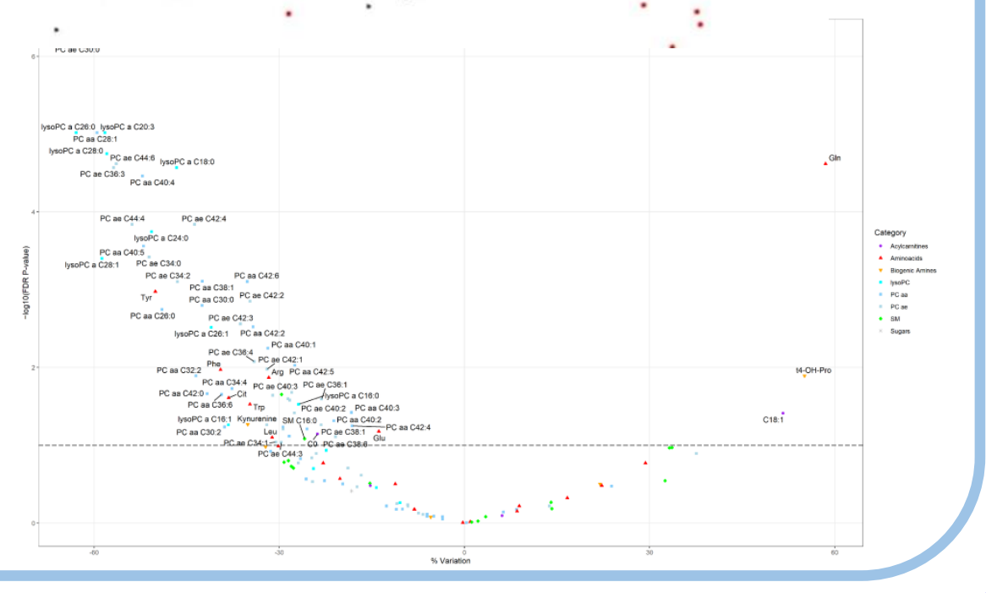
| | |
|---|---|
| gentab_lm_long | Given a dataframe and a set of dependent and independent variables from that dataframe, it generates a linear regression model for each single dependent variable and creates a new table with all slopes and *p*-values. |

- Linear models with fixed effects
- Linear model with mixed effect (random and fixed)
- Linear TOBIT model, for considering non quantifiable data as left-censored



| | |
|---|---|
| getBoxplots | Given a table containing data, it creates boxplots for each desired variable. |
| Volcano_lm | Given a table generated with the function gentab_lm_long, it creates a Volcano plot. |





## Conclusions

In conclusion, the GetFeatistics is particularly helpful for streamlining the metabolomics data elaboration of epidemiological studies within the R environment. The functions have been specifically developed for being applied to high numbers of features such as for metabolomics studies. The strategy to process pooled quality controls and the linear model with mixed effects represent the main advantages of this package.

## Acknowledgment

I acknowledge Albina Rastoder, who helped with the testing of some of the functions

The package is on GitHub: https://github.com/FrigerioGianfranco/GetFeatistics

**20th Annual Conference of the Metabolomics Society – METABOLOMICS 2024 – Osaka, Japan, 16th -20th June 2024**