Streamlined advanced statistics in Metabolomics: the GetFeatistics R package

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

Gianfranco Frigerio
Any questions? Write me!!! ☺
frigerio.gianfranco@hsr.it
🌐 www.gianfrancofrigerio.info
in Gianfranco Frigerio
@gianfrancofrigerio.bsky.social

I.R.C.C.S. Ospedale San Raffaele — Gruppo San Donato
COSR   ProMeFa
UNIVERSITÉ DU LUXEMBOURG
Luxembourg Centre for Systems Biomedicine

Why this package?

Metabolomics studies often rely on tools like MS-DIAL and R-based algorithms (e.g., XCMS) to process raw data into feature tables. Previous research has shown that using separate pooled QC samples can enhance dataset quality while preserving low-concentration metabolites, such as exogenous compounds. In observational epidemiologic studies, controlling for confounders is essential for accurate metabolite association analysis. For longitudinal studies, linear mixed-effects models provide a robust statistical approach. The purpose of this package is to streamline such elaboration of metabolomics data analysis.
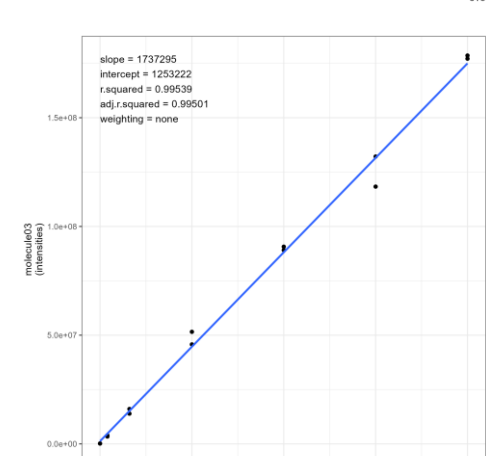
Workflow

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

Non-targeted analyses

get_feat_info_from_MSDial, **get_feat_table_from_MSDial**, **get_feat_info_from_patRoon**, **get_feat_table_from_patRoon** — From MS-DIAL output or from patRoon output, it creates the "featTable" and the "featINFO". Given pre-defined cut-off values, it also assigns annotation levels to each feature to featINFO.

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

checkmolecules_in_feat_table — Function to monitor internal standards (or native standards) in the table of features

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

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 suitable for following elaborations and statistical analyses (samples in rows and features in columns)

Statistical analyses

test_normality_density_plot, **test_normality_q_q_plot**, **test_normality_saphiro_table** — Functions to test normality, by producing density plot, q-q-plot and performing Shapiro test

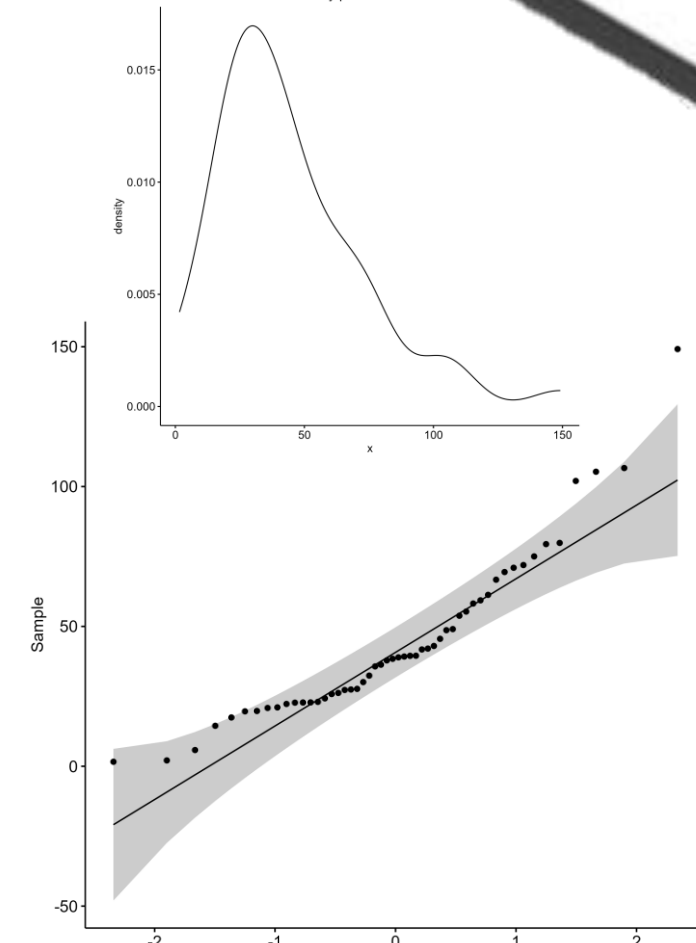**gentab_descr** — generate a table with descriptive statistics for each of those variables

transf_data — To replace missing values, log-transform, and/or scale the data. (auto_scale, pareto_scale, mean_scale, range_scale)

clean_transf_colnames — To keep only transformed data

Univariate analyses

gentab_P.t.test — It performs t-test to each desired variable and creates a new table with the p-values

gentab_FC — It performs Fold Change analysis, taking into account whether variables were log-transformed

Volcano_ttest_FC — Combines t-test and Fold Change in a graph

gentab_P.1wayANOVA_posthocTukeyHSD — It performs one-way ANOVA, with also TukeyHSD posthoc tests for between-groups comparison, to each desired variable and creates a new table with the p-values

gentab_FC_more_than2levels — It performs Fold Change analysis for each pair of groups

gentab_P.2wayANOVA_posthocTukeyHSD — It performs 2-way ANOVA, with also TukeyHSD posthoc tests for between groups comparison, to each desired variable and creates a new table with the p-values
- Interactions test can also be performed

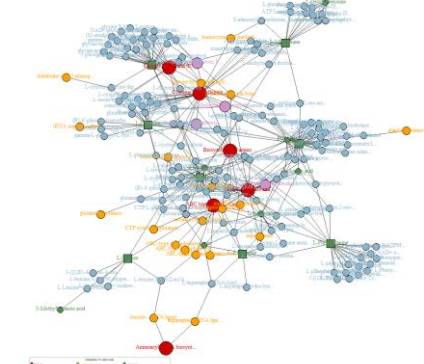**getBoxplots** — Given a table containing data, it creates boxplots for each desired variable

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

Multivariate analyses

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
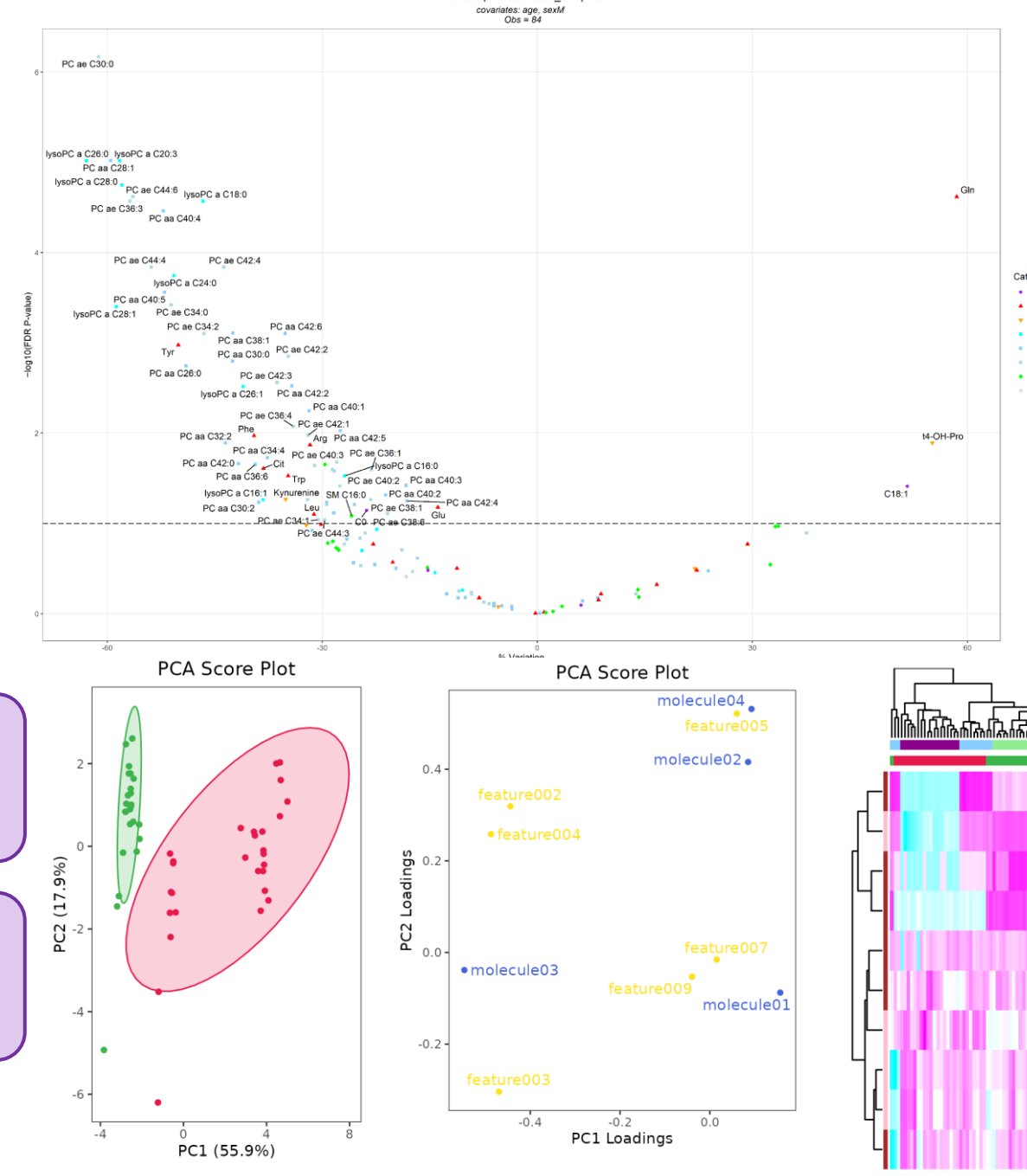
Volcano_lm — Given a table generated with the function gentab_lm_long, it creates a Volcano plot

getPCA — Build tables and plots with scores and loadings of a Principal Component Analysis

getHeatMap — Build a heat map graph, with the option of grouping and cluster analyses.

Pathway Enrichment Analysis

do_FELLA_enrichment_analysis — A pathway enrichment analysis is performed with the FELLA package

Miscellaneous

fix_names — It removes special characters from a character vector

merge_results — To merge the results from different chromatographic run/polarities into a single table

get_phase_amount — To know the exact amount mobile phases needed for a gradient

Conclusions

In conclusion, the GetFeatistics R-package streamlines metabolomics data processing for epidemiological studies in R. It is designed to handle large feature sets, with key advantages including pooled QC processing and linear mixed-effects modeling.

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