

## Standard Procedure For Analysis

**Project Name:** General Metabolomics Analysis SOP

**Requestor:** Analysis Request Sent by (Name and Contact info)

**Names and contact info of Analyst:**

**Brief Description of Analysis Types:** Data are normalized, examined for quality, and analyzed for significant difference between experimental groups.

**Procedure or Analysis Revision Date:** 2/19/13

**Analysis Results/Deliverable Date:**

---

**Scientific context or hypothesis for analysis:**

**Checklist before starting analysis:**

**Sources of input required:**

Such as:

- Experiment design,
- Sample annotation,
- Species,
- Tissue and other supporting information

**Input data:**

- Tab delineated text file created from median corrected “scaled” worksheet of Metabolon output of MS metabolite measures
- Downloaded from sysbiocube

**Analysis plan or steps taken:**

- Data format expected:
  - .tsv file of median normalized metabolomics data
  - One metabolite per row (mappable to KEGG id)
  - One sample per column named in such a way that experimental groupings are clear (mappable to patient or animal data)
  - A separate .tsv file may contain information on metabolite groupings, pathways, etc.

- Data preparation (Metabolon output and R)
  - If not already done, rows will be median normalized and missing values imputed from minimum values.
  - Sample (column) normalization:
    - For exploratory metabolomics screens (not targeted), the distribution of row means will be analyzed for uniformity and outliers and normalized if necessary.
  - Upload normalized .tsv to SysBioCube
  - Cluster sample groups by metabolite values
    - Upload to SysBioCube
- DEM identification (all done in R).
  - unpaired t-test, unequal-variance, significance level 0.05
  - Benjamini and Hochberg multiple comparison correction
  - Upload DEM list and associated statistics to SysBioCube
- DEM confirmation/testing
  - Permutation tests
  - ROC curves
- KEGG pathway analysis
  -

**Statistical Analysis Plan or procedure:**

**Parameters Used:**

**Generation of research questions:**

**Software Used:**

- Such as gene spring, R packages, Partek and etc.

**Software/program/script developed:**

**Databases and public data sources:**

- Such as Kegg pathways, DAVID, biocarta, HMDB or etc..

**Data Disposal:**

- This will include where analysis files, results are saved at common location at Sysbiocube (upload)
- File names including intermediate files

**Short Description of results or finding:**

- Such as list of biomarkers, figures

**Publications and references:**

**Analysis Tasks performed:**

- Analysis steps performed by analyst (Name and Task)