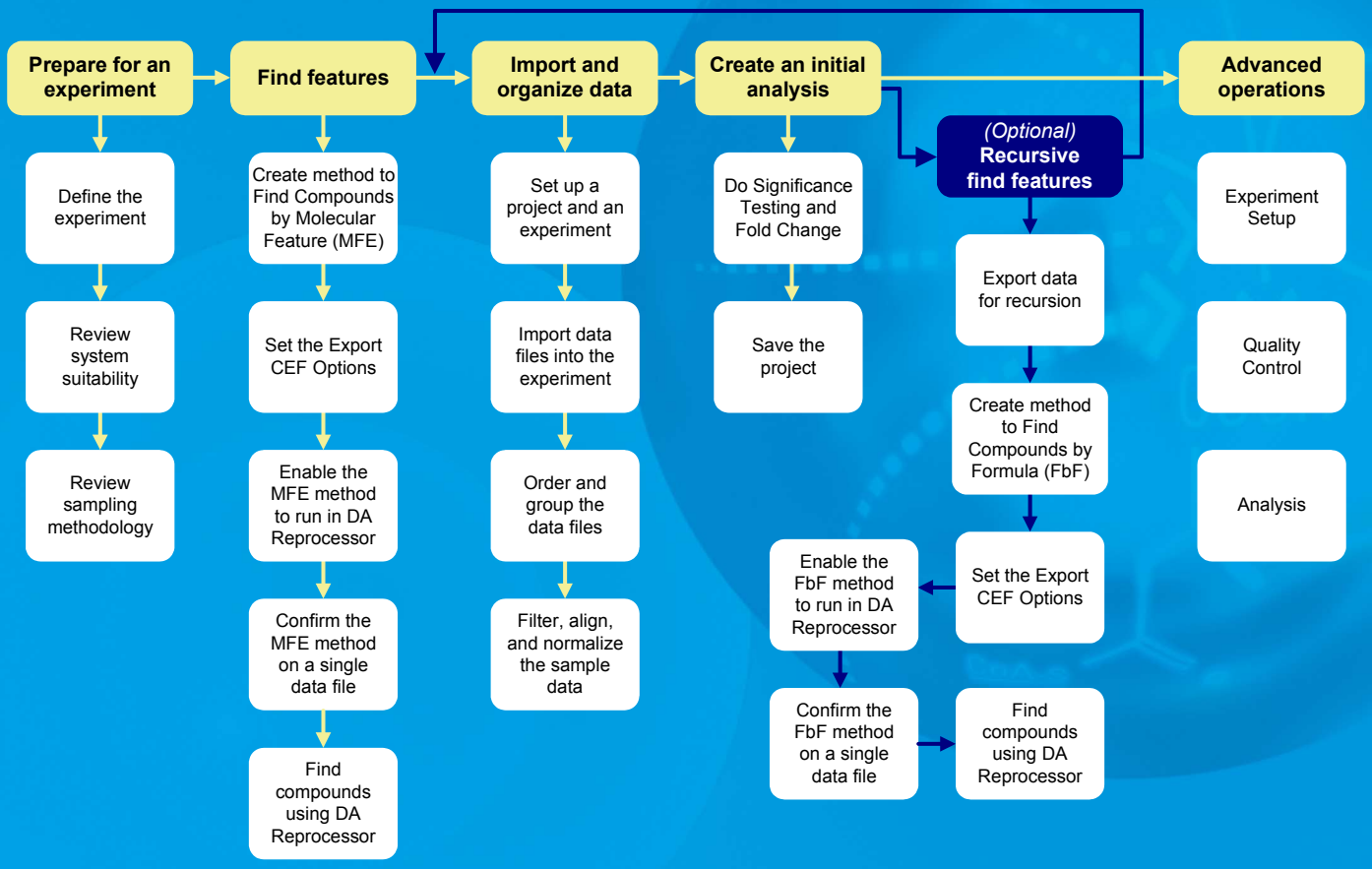
Agilent Metabolomics Workflow

Discovery Workflow Overview



Agilent Technologies

Introduction to Metabolomics

Metabolomics is an emerging field of 'omics' research that is concerned with the characterization and identification of metabolites, the end products of cellular metabolism. Metabolomics research leads to complex data sets involving hundreds to thousands of metabolites. Comprehensive analysis of metabolomics data requires an analytical approach and data analysis strategy that are often unique and require specialized data analysis software that enables cheminformatics analysis, bioinformatics, and statistics. Agilent products provide you with the necessary data analysis tools.

The Agilent Metabolomics Workflow guides you through the use of Agilent MassHunter Qualitative Analysis and Agilent Mass Profiler Professional to set up your experiment design, collect replicate samples, assess system suitability, improve data quality, and perform comprehensive analysis of metabolomics data.

MassHunter Qualitative Analysis software allows you to automatically find and extract all spectral and chromatographic information from a sample, even when the components are not fully resolved. Powerful data navigation capabilities permit you to browse through compound-specific information in a single sample and compare chromatograms and spectra among multiple samples. The software also includes a customizable user interface and the capability to save, export or copy results into other applications, such as Mass Profiler Professional.

Mass Profiler Professional is a powerful data reduction, analysis, and visualization tool that allows you to identify statistically significant answers to simple questions presented to complex data sets; data sets such as the chemical signatures from biological matrices encountered in metabolomics.

You can use the Agilent Metabolomics Workflow as a road map for any analysis that requires the identification of statistically significant answers to simple questions presented to complex data sets. The metabolomics workflow may be used to perform the following analyses:

- Compare two or more biological groups
- Find and identify potential biomarkers
- Look for biomarkers of toxicology
- Understand biological pathways
- Discover new metabolites
- Develop data mining and data processing procedures that produce characteristic markers for a set of samples
- Construct statistical models for sample classification.

Small molecule tuning (LC/MS)

Metabolomics involves the analysis of small molecules, molecules nominally with a molecular weight from 50 to 6000 amu. Since the best results for metabolomics studies involve the identification of the exact mass of the molecular ion, LC/MS instrument tunes should be adjusted to (1) improve the sensitivity for intact molecular ions and (2) improve the overall sensitivity for small, low-mass, compounds. A typical automated instrument tune optimizes the instrument sensitivity across the entire mass range and may result in lower sensitivity for small molecules combined with higher fragmentation than desired for metabolomics.

Required Items

The Metabolomics Workflow performs best when using the hardware and software described in the "required" sections below. The required hardware and software is used to perform the data analysis tasks shown in Figure 1.

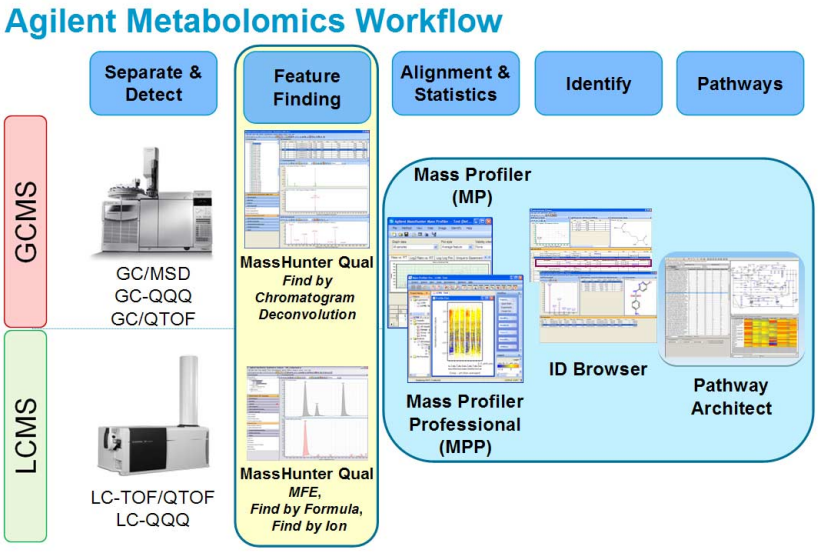


Figure 1 *Agilent hardware and software used in performing metabolomics.*

A typical Agilent metabolomics workflow is illustrated in Figure 2 starting with data acquisition through to analysis involving both untargeted (discovery) LC/MS and targeted (confirmation) LC/MS/MS analyses. Molecular feature extraction (MFE) and Find by Formula (FbF) are two different algorithms used by MassHunter Qualitative Analysis for finding compounds. All results files generated by Agilent analytical platforms can be imported into Mass Profiler Professional for quality control, statistical analysis and visualization, and interpretation.

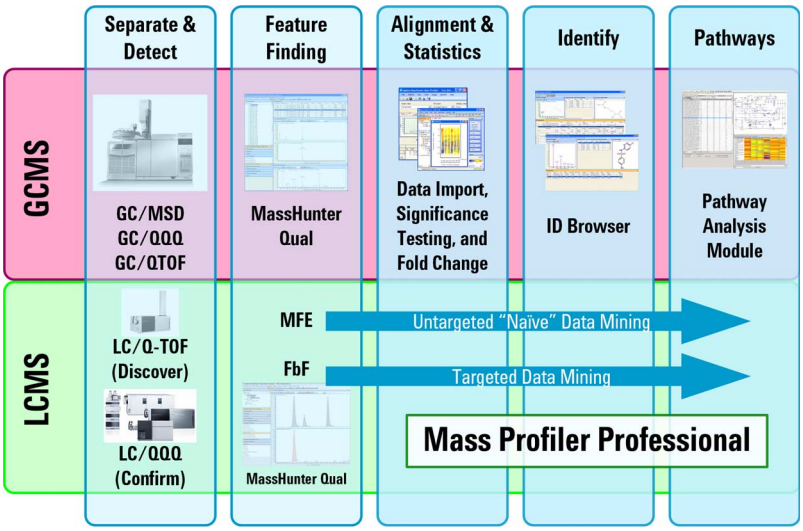


Figure 2 *An Agilent metabolomics workflow from separation to pathway analysis typically involving either or both GC/MS and LC/MS analyses.*

Required hardware

- PC running Windows
  - *Minimum:* XP SP3 (32-bit) or Windows 7 (32-bit or 64-bit) with 4 GB of RAM
  - *Recommended:* Windows 7 (64-bit) with 8 GB or more of RAM
- At least 50 GB of free space on the C Partition of the hard drive
- Data from an Agilent GC/MS, LC/MS, CE/MS and/or ICP-MS system or data that may be imported from another instrument.

Required software

- Agilent Mass Profiler Professional Software B.12.00 or later
- Agilent MassHunter Qualitative Analysis software, Version B.03.01, B.04.00, B.05.00 SP1 or later
- Agilent MassHunter Data Acquisition software, Version B.03.02, B.04.00, B.05.00 or later (this will include Agilent MassHunter DA Reprocessor)
- Agilent MassHunter Quantitative Analysis software, Version B.03.02 or later

Optional software

- Agilent ChemStation software
- AMDIS
- MassHunter ID Browser B.03.01 or later
- METLIN Personal Compound Database and Library
- Agilent Fiehn GC/MS Metabolomics Library

What is metabolomics?

Metabolomics is the process of identifying and quantifying all metabolites of an organism in a specified biological state. The metabolites of an organism represent a chemical "fingerprint" of the organism in a well defined state where the state is defined by a set of specific circumstances.

The metabolomics workflow involves the study of two types of variables:

Independent variables: When one or more of the attributes of the state of the organism are known to you in advance of sampling. These attributes are referred to as an independent variable.

During the various steps of the metabolomic data analysis the metabolomics workflow refers to the known states of the organism, or externalities to which the organism is subjected, as parameter values, conditions, or attribute values. The known states and externalities represent independent variables in the statistical analyses.

Dependent variables: The observable biological response to changes in the independent variables. The response can be manifest in a change in the metabolic profile. Each metabolite that undergoes a change in expressed concentration is referred to as a dependent variable.

The metabolites in a sample may be individually referred to as compounds, features, elements, or entities during the various steps of the metabolomic data analysis. Metabolites represent dependent variables in the statistical analyses.

When hundreds to thousands of metabolites exist, researchers employ chemometric data analyses to reveal accurate and statistically meaningful correlations between the independent variables and the metabolic profile. They use meaningful information learned from the metabolite responses for clinical diagnostics, for understanding the onset and progression of human diseases, and for treatment. This analysis is the metabolomics workflow illustrated in Figure 3.

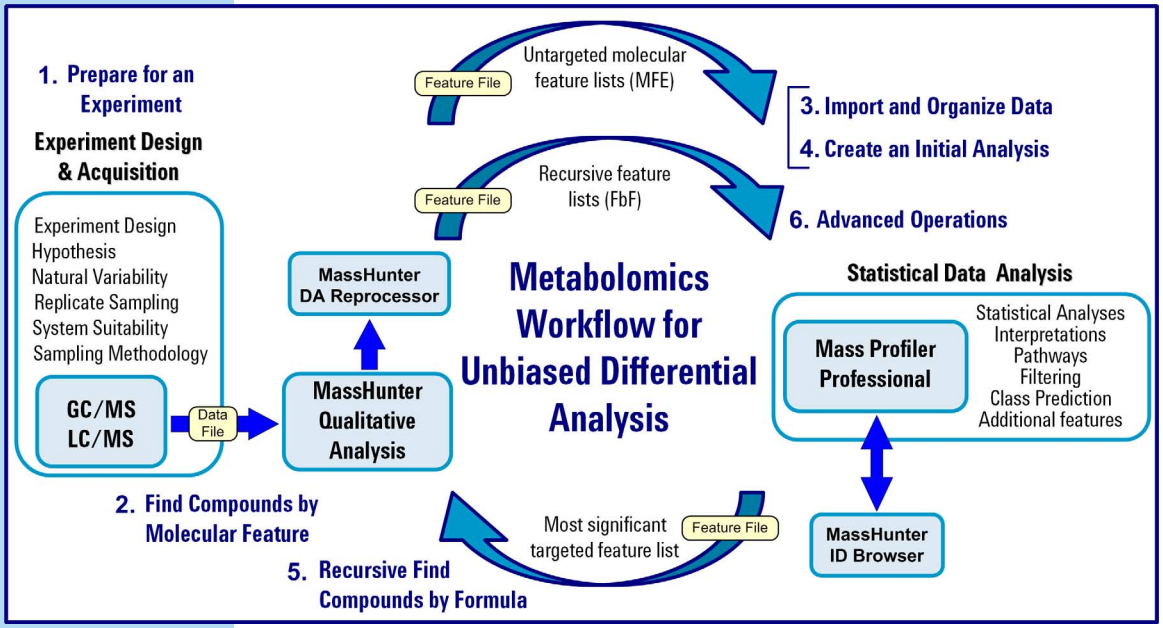


Figure 3 *Agilent Metabolomics Workflow flow diagram*

The hypothesis

The first and most important step in the metabolomic process is to formulate the question of correlation that is answered by the analysis - the hypothesis. This question is a statement that proposes a possible correlation, for example a cause and effect, between a set of independent variables and the resulting metabolic profile. The metabolomic workflow is used to prove or disprove the hypothesis.

Natural variability

Before the metabolomics workflow is begun it is important to understand how any one sample represents the population as a whole. Because of natural variability and the uncertainties associated with both the measurement and the population, no assurance exists that any single sample from a population represents the mean of the population. Thus, increasing the sample size greatly improves the accuracy of the sample set in describing the characteristics of the population.

Replicate sampling

Sampling the entire population is not typically feasible because of constraints imposed by time, resources, and finances. On the other hand, fewer samples increase the probability of concluding a false positive or false negative correlation. At a minimum, it is recommended that your metabolomics analysis include ten (10) or more replicate samples for each attribute value for each condition in your metabolomics study.

System suitability

System suitability involves collecting data to provide you a means to evaluate and compensate for drift and instrumental variations to assure quality results. The techniques employed by the metabolomics workflow of (1) retention time alignment, (2) intensity normalization, (3) chromatographic deconvolution, and (4) baselining produce the highest quality results. However, the process cannot compensate for excessive drift in the acquisition parameters. The best results are achieved by maintaining your instrument and using good chromatography.

Sampling methodology

Improved data quality for metabolomics analysis comes from matching the sampling methodology to the experimental design so that replicate data is collected to span the attribute values for each condition. A larger number of samples appropriate to the population under study results in a better answer to the hypothesis. An understanding of the methodologies used in sampling and using more than one method of sample collection have a positive impact on the significance of the results.

Introduction to the workflow

The Agilent metabolomics workflow consists of six step and involves the use of Agilent MassHunter Qualitative Analysis, Agilent MassHunter DA Reprocessor, and Agilent Mass Profiler Professional.

Step 1

Prepare for an Experiment

The first and most important step in the metabolomics workflow is to formulate your hypothesis, the question of correlation that is answered by the analysis. Your preparation includes an experiment definition that considers natural variability and replicate sampling, and reviews system suitability and sampling methodology to improve the significance of your analysis.

Step 2

Finding Features

Find Compounds by Molecular Feature (MassHunter Qualitative Analysis)
Compounds, referred to as molecular features, are extracted from your data based on mass spectral and chromatographic characteristics. The process is referred to as Molecular Feature Extraction (MFE). Molecular feature extraction quickly and automatically generates a complete, accurate list of your compounds which include molecular weight, retention time, m/z, and abundance.

Step 3

Import and Organize Data

Organize, import, and prepare your data (Mass Profiler Professional)
After you create a project and an experiment, the "MS Experiment Creation" guides you through the necessary steps to organize your experiment, import your data, define your experimental variables, and prepare your data for analysis. The data preparation includes filtering, alignment, normalization, and baselining.

Step 4

Create an Initial Analysis

Quality control and initial differential expression (Mass Profiler Professional)
The "Significance Testing and Fold Change Wizard" guides you through the necessary steps to enter parameters and values that improve the quality of your results and produce an initial differential expression for your analysis.

Step 5

Recursive Find Features *(Optional)*

Find Compounds by Formula (MassHunter Qualitative Analysis)
Importing the most significant features back into MassHunter Qualitative Analysis as targeted features improves finding the features in your samples. This repeated feature finding is referred to as recursion. Improved reliability in finding your features leads to improvement in the accuracy of your analysis.

Step 6

Advanced Operations

Customize your analysis and interpret the results (Mass Profiler Professional)
The most significant features in your data are processed by Mass Profiler Professional into a final statistical analysis and interpretation. The results from the final interpretation may be used to prove or disprove your hypothesis and may be used to create a sample class prediction model.

Working with Mass Profiler Professional

Mass Profiler Professional imports CEF files created by MassHunter Qualitative Analysis. Mass Profiler Professional exploits the high information content of chromatography/mass spectrometry data and can easily import, analyze and visualize GC/MS, LC/MS, CE/MS and ICP-MS data from large sample sets and complex MS data sets.

Mass Profiler Professional provides you with a step-by-step set of guides to (1) set up a project and an experiment, (2) import the data files, (3) order and group the data files, (4) filter, align, and normalize the sample data, and (5) create an initial analysis.

Because the advanced operations available in the Workflow Browser do not guide you through the initial steps of data import and differential analysis, it is not recommended to skip the "Import and organize data" or "Create an initial analysis" steps of the metabolomics workflow. All parameters, including the default parameters used during the MS Experiment Creation Wizard, can be edited at the conclusion of the Metabolomics Workflow by using the operations available in the Workflow Browser (Figure 4 on page 12).

Note: Help and detail information regarding the various fields and statistical treatments are available at anytime by pressing **F1** or referring to the *Mass Profiler Professional User Manual*.

Import and organize your data

After you create a project and an experiment, the "MS Experiment Creation Wizard" guides you through the necessary steps to organize your experiment, import your data, define your experimental variables, and prepare your data for analysis. There are four main steps to create your experiment: (1) set up a project and an experiment, (2) import the data files, (3) order and group the data files, and (4) filter, align, and normalize the sample data.

a Start Mass Profiler Professional

Double-click the Mass Profiler Professional icon () located on the desktop, or
Click **Start > All Programs > Agilent > MassHunter Workstation > Mass Profiler Professional > Mass Profiler Professional**

b Set up a project and an experiment

A project is a container for a collection of experiments. A project can have multiple experiments on different sample types and organisms. Four steps are involved in the project setup:

Startup: Click **Create new project**.

Create New Project: Type descriptive information about the project.

Experiment Selection Dialog: Click either **Create new experiment** or **Open existing experiment**.

New Experiment: Type and select information that guides the experiment creation.

c Import your data files into the experiment.

Your data files are imported in to Mass Profiler Professional during Step 1 and Step 2 of the MS Experiment Creation wizard.

Step 1. Select Data Source: Select the data source that generated the molecular features you are using for your experiment.

Step 2. Select Data to Import: Select the molecular feature sample files.

d Order and group your data files.

Your data files are ordered in Step 5 and your experimental grouping is entered in Step 6 of the MS Experiment Creation wizard.

Step 5. Sample Reordering: Organize your samples by selecting and deselecting individual samples and reordering the selection to group the samples based on the independent variables.

Step 6. Experiment Grouping: Define the sample grouping with respect to your independent variables, including the replicate structure of your experiment.

e Filter, align, and normalize the sample data.

You filter, align, and normalize your sample data in Step 7 through 11 of the MS Experiment Creation wizard. At each step of the process you can view your progress and return to prior steps to adjust your results.

Step 7. Filtering: Filter the molecular features by abundance, mass range, number of ions per feature, and charge state.

Step 8. Alignment: Align the features across the samples based on tolerances established by retention time and mass. This step is omitted when the experiment type is "identified" because identified compounds are treated as aligned by identification.

Step 9. Sample Summary: Display a mass versus retention time plot, spreadsheet, and compound frequency for the distribution of aligned and unaligned entities in the samples. Compound Frequency charts provide a quick view into the effectiveness of the alignment of unidentified experiment types. The back and next buttons in the wizard let you easily review the effects of different alignment and filter options.

Step 10. Normalization Criteria: Scale the signal intensity of sample features to a value calculated by the specified algorithm or an external scalar.

Step 11. Baselining Options: Compare the signal intensity of each sample to a representative value calculated across all of the samples or the control samples.

You have now completed the import and organization of your data. In the next workflow step you create an initial differential expression from your data using Mass Profiler Professional.

Create an initial analysis

The "Significance Testing and Fold Change Wizard" guides you through the necessary steps to enter parameters and values that improve the quality of your results and produce an initial differential expression for your analysis.

The Significance Testing and Fold Change workflow helps you create an initial differential expression from your data and identify the most significant features from among all of the features previously found using molecular feature extraction. The steps necessary to create your initial differential expression are predetermined and based on the experiment type, experiment grouping, and conditions you entered when creating your project and setting up your experiment.

The workflow displays the sequence of steps on the left-hand side navigator with the current step highlighted. Some steps may be automatically skipped for your experiment. All of the parameters can be edited at the conclusion of the Analysis: Significance Testing and Fold Change workflow by using the operations available in the Workflow Browser.

Step 1. Summary Report: Displays a summary view of your experiment based on the parameters you provided in the Import Data wizard. A profile plot with the samples on the x-axis and the log normalized abundance values on the y-axis is displayed. If the number of samples is more than 30, the data is represented by a spreadsheet view instead of a profile plot.

Step 2. Experiment Grouping: Independent variables and the attribute values of the independent variables must be specified to define grouping of the samples. An independent variable is referred to as a parameter name. The attribute values within an independent variable are referred to as parameter values. Samples with the same parameter values within a parameter name are treated as replicates.

Step 3. Filter Flags: The compounds created during the experiment creation are now referred to as entities. The entities are filtered (removed) from further analysis based on their presence across samples and parameter values (now referred to as a condition).

Step 4. Filter by Frequency: Entities are further filtered based on their frequency of presence in specified samples and conditions. This filter removes irreproducible entities.

Step 5. Quality Control on Samples: The samples are presented by grouping and the current Principal Component Analysis (PCA). PCA calculates all the possible principal components and visually represents them in a 3D scatter plot. The scores shown by the axes scales are used to check data quality. The scatter plot shows one point per sample colored-coded by the experiment grouping. Replicates within a group should cluster together and be separated from samples in other groups

Step 6. Significance Analysis: The entities are filtered based on their p-values calculated from a statistical analysis. The statistical analysis performed depends on the samples and experiment grouping.

Step 7. Fold Change: Compounds are further filtered based on their abundance ratios or differences between a treatment and a control that are greater than a specified cut-off or threshold value.

Step 8. ID Browser Identification: The final entity list is directly imported into ID Browser for identification and returned to Mass Profiler Professional.

The Significance Testing and Fold Change workflow lets you to proceed through each step using the **Next** button. A summary of your analysis is presented in each subsequent step. After review of your analysis progress you may return to any previous step and make changes by using the **Back** button. To become more familiar with the analysis parameters and how the parameters affect your data it is recommended that you frequently use the **Back** and **Next** buttons.

To exit the wizard and skip the later steps in the wizard, click **Finish** at any step. When you click **Finish**, the entity list is saved and you may commence analysis using the advanced operations available in the Workflow Browser.

Advanced operations

The most significant features in your data are processed by Mass Profiler Professional into a final statistical analysis and interpretation. The results from the final interpretation may be used to prove or disprove your hypothesis and may be used to create a sample class prediction model or evaluate pathways.

The operations available in the Workflow Browser provide you with the tools necessary to analyze features from your mass spectrometry data depending upon the need and aim of your analysis, the experimental design, and the focus of the study. This helps you to create different interpretations to carry out the analysis based on the different filtering, normalization, and standard statistical methods.

Because the advanced operations available in the Workflow Browser do not guide you through the initial steps of data import and differential analysis, it is not recommended to skip the “Import and organize data” or “Create an initial analysis” steps of the metabolomics workflow. All parameters, including the default parameters used during the MS Experiment Creation Wizard, can be edited at the conclusion of the Metabolomics Workflow by using the operations available in the Workflow Browser (Figure 4 on page 12).

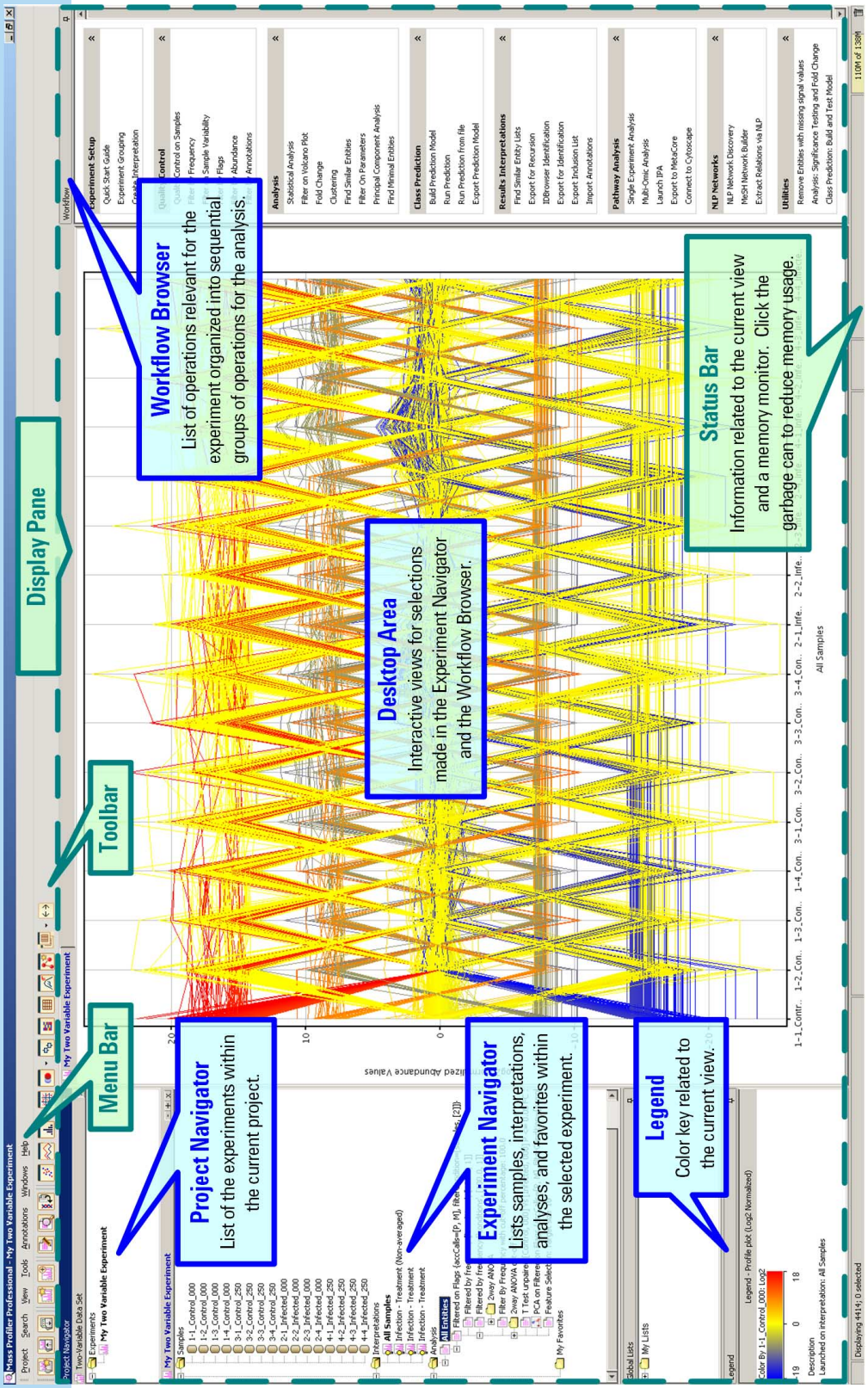


Figure 4 *The main functional areas of Mass Profiler Professional*

For More Information

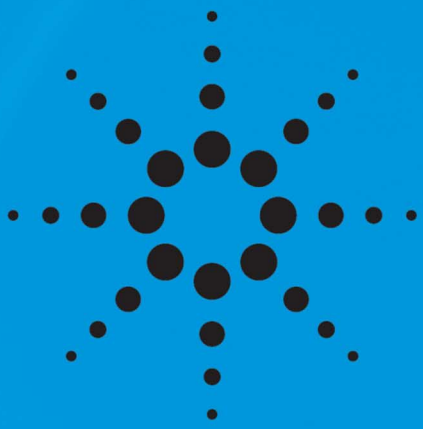
For details about these procedures, see the Agilent Metabolomics Discovery Workflow Guide. Visit our Web site at www.agilent.com/chem.

The metabolomics discovery workflow is part of the collection of Agilent manuals, help, application notes, and training videos. The current collection of manuals and help are valuable to users who understand the metabolomics workflow and who may require familiarization with the Agilent software tools. Training videos provide step-by-step instructions for using the software tools to reduce example GC/MS and LC/MS data but require a significant time investment and ability to extrapolate the example processes. This workflow provides a step-by-step overview of performing metabolomics data analysis using Agilent MassHunter Qualitative Analysis and Agilent Mass Profiler Professional.

The following selection of publications provides materials related to metabolomics and Agilent Mass Profiler Professional:

- *Manual:* Agilent Metabolomics Workflow - Discovery Workflow Guide (Agilent publication 5990-7067EN, Revision B, October 2012)
- *Brochure:* Agilent Solutions for Metabolomics (Agilent publication 5990-6048EN, April 30, 2012)
- *Brochure:* Agilent Mass Profiler Professional Software (Agilent publication 5990-4164EN, April 27, 2012)
- *Application:* Mass Profiler Professional and Personal Compound Database and Library Software Facilitate Compound Identification for Profiling of the Yeast Metabolome (Agilent publication 5990-9858EN, April 25, 2012)
- *Application:* Multi-omic Analysis with Agilent's GeneSpring 11.5 Analysis Platform (Agilent publication 5990-7505EN, March 25, 2011)
- *Presentation:* Multi-omic Analysis Software for Targeted Identification of Key Biological Pathways (Agilent publication USHUPO_IB_March_2012.pdf, March 2012)
- *Application:* An LC/MS Metabolomics Discovery Workflow for Malaria-Infected Red Blood Cells Using Mass Profiler Professional Software and LC-Triple Quadrupole MRM Confirmation (Agilent publication 5990-6790EN, November 19, 2010)
- *Brochure:* Integrated Biology from Agilent: The Future is Emerging (Agilent publication 5990-6047EN, September 1, 2010)
- *Primer:* Metabolomics: Approaches Using Mass Spectrometry (Agilent publication 5990-4314EN, October 27, 2009)

A complete list of references may be found in "References" in the Agilent Metabolomics Workflow.



www.agilent.com

Revision B, October 2012

5990-7068EN

Agilent Technologies