Evaluating Metabolomics Data with Ideom

This help file is intended for visualisation and evaluation of data that has already been processed with Ideom.

1. Open the file in Microsoft Excel version 2007 or 2010 (not Mac versions) and enable macros from the warning at the top of the screen.
2. The Ideom file contains a lot of information, but the summary on the 'comparison' sheet should be sufficient for most biochemical interpretation.
3. The data is presented as a list of detected peaks, and each peak has been putatively identified by mass and retention time. Metabolites highlighted yellow (with a confidence score of 10) are confirmed identities based on authentic standards, all other metabolites are putative identifications based on our database of 40,000+ possible metabolites, with varying levels of confidence as shown.
4. Metabolite levels for each experimental group are shown in columns J onwards. Levels are expressed as mean peak intensity (height) relative to the mean peak intensity of the Control group. Additional statistical data for each group is recorded in additional columns to the right hand side. NOTE: The Control group is defined in the 'settings' sheet. Groups may be re-assigned or excluded from analysis in the settings sheet (column H), however for changes to take effect on the Comparison sheet you need to re-run step 8 (Green button: 'Compare all sets').
5. Metabolite levels are coloured according to relative intensity (blue = low, red = high) and highlighted in bold type if the p-value (from t-test against Control group) is less than 0.05.
6. The simplest way to evaluate the data is to sort metabolites according to relative intensities in a specific experimental group (either with the orange Sort button or with the native Excel sort function). This will group those metabolites that change the most for a particular experimental condition. Alternatively, metabolites that follow a certain trend across multiple experimental groups (eg. for a time-course) can be grouped by using the 'Correlation Sort' button.
7. Use the native Excel Filter function to focus on specific metabolites of interest. For example you may wish view only metabolites with high identity confidence (column F), or exclude peptides and lipids (column G), or only those with significant p-values.
8. Double-click individual metabolite names to get a column chart of raw peak intensities with standard deviations.
9. Use the orange buttons to plot graphs for visualisation of multiple metabolites (NOTE: you may wish to apply filters before plotting some graphs to avoid over-complication).
10. Multivariate statistics can be undertaken by the light blue 'Export to R or Metaboanalyst' button. The R option will automatically generate a heat map, HCA dendrograms, PCA plot and csv file with the PCA loadings. The Metaboanalyst option generates a file in the correct format to upload to this user-friendly website to do your own analysis.
11. Biochemical pathway interpretation can be done by export to Pathos (for KEGG maps) and Metexplore (for MetaCyc networks). Alternatively, sorting (or filtering) by Map (column G) and Pathway (column H) will give a good summary of metabolite levels for specific pathways.

Caution: Metabolite identification

- Metabolites identified by mass and retention time (with authentic standards) are highlighted yellow and given a confidence score of 10, all others are putative identifications.

- Putative identification is by an untargeted method whereby high-resolution LCMS signals are matched to a database of theoretical masses and predicted retention times. LCMS data contains a lot of noise, and Ideom processing automatically removes as much as possible, but some false identifications will remain.

- The most common false-identifications are MS in-source artefacts (fragments and adducts). These can be identified by checking the mass differences to co-eluting compounds (double-click in column A for spectra or column F for annotations, consider sorting the list by retention time to identify these).

- Other noise peaks may be identified by looking at the reproducibility of peak intensities across replicate samples (double-click in column I) or assessment of the LCMS peak shape (additional files required).

- Isomeric peaks (formula shown in red text) may be due to genuine isomeric metabolites (real), or poor chromatography of a single metabolite (false). Check the RT and intensity of isomeric peaks by double-click in column B (consider sorting the list by mass or formula to group isomeric peaks).

- Putative identification is often difficult in cases where numerous putative isomers exist in the database for a single mass (formula). The most likely isomer is identified by a complex algorithm, however you should always check the list of alternative isomers that could correspond to each mass (double-click in column D). Alternative isomers can be manually selected from the dropdown list that is activated by a single-click of the metabolite name.

- Unidentified peaks (or unusual identities) can be investigated by double-click in column C. The yellow box lists possible identifications (with ppm error) for common adducts. You also have the option to search for a formula (that is not in our database) based on the exact mass (using either rCDK, Xcalibur or Chemspider).

- Please confirm the identity of any ‘putative identities’ with your LCMS specialist. Occasionally interesting-looking peaks can be excluded as noise or artefacts after looking at the raw data.

> Example 1: Cidofivir (mass=279.062; RT=18.24) levels appear to change under certain experimental conditions, yet the microorganism under investigation has not been treated with this antiviral drug. Closer inspection reveals a similar pattern of peak intensities for cytidine, which also elutes RT=18.24 mins, and the mass difference of 35.977 corresponds to chlorine. Therefore the ‘cidofivir’ is actually an in-source chlorine adduct of cytidine.
>
> Example 2: UDP appears twice in a list, so which one should you take? Looking at the retention times and peak intensities we can conclude that the peak at 18.22 min is probably an in-source fragment of UDP-glucose (RT=18.22), the peak at 18.44 min is unique and therefore probably the real UDP peak.