nonlinear DYNAMICS

Progenesis CoMet

Concepts and challenges of Metabolomics data analysis.

Mark Bennett – General Manager

24th July, 2013

Background of Nonlinear Dynamics

- We develop innovative high quality data analysis software for proteomics and metabolomics to help you:
  - Achieve breakthroughs and increase scientific understanding
  - Make real discoveries that you can pursue with confidence, leading to improvements in life quality for everyone
- Founded 1989 with head office in Newcastle upon Tyne, UK
  - US office based in Raleigh Durham, North Carolina

nonlinear DYNAMICS

Progenesis™ Concepts

1. Based on a common concept, solving the key challenges in relative quantification
   - Progenesis LC-MS & Progenesis CoMet also benefit from a quantify-then-identify approach
2. Developed with key opinion leaders within proteomics and metabolomics
3. Adopted and proven across many labs around the world

Products:

- Progenesis SameSpots for 2D gel based proteomics
- Progenesis LC-MS for MS-based proteomics
- Progenesis CoMet for MS-based metabolomics
- Progenesis MALDI for MS-based proteomics

nonlinear DYNAMICS

Goal of metabolomics discovery?

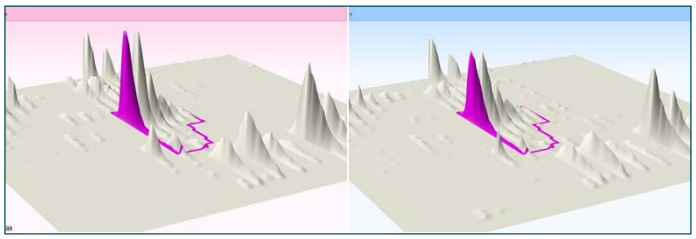
"To identify the **compounds** that warrant further investigation as **rapidly**, **objectively** and **reliably** as possible."

- **Progenesis CoMet** developed with this in mind...
  - Produce a comprehensive **table of detected compounds**, which you can easily share or validate and put them in biological context
  - A complete analysis approach to combine quantification and identification of significantly changing compounds, in **one streamlined package**
  - An **objective approach** to analysis, using a complete matrix of data with no missing values, for results based on **reliable statistics**

nonlinear DYNAMICS

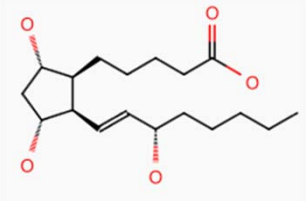
Relative quantification of metabolites

- LC-MS (ESI) is one technique used for such quantification
- Examples of applications:
  - **Discovery** - relative changes in metabolite abundance related to experimental conditions
  - **Targeted** – monitoring the relative abundances of numerous known compounds simultaneously



Key steps in metabolomics discovery

- Relative quantification of metabolites
  - RT data alignment
  - Peak picking including **ion deconvolution**
  - Quality control, particularly of LC and quant
  - Data normalisation
  - Statistical analysis



- Identification
  - Several approaches (e.g. m/z, neutral mass, ± RT, ± MS/MS spectra)
  - No standardised search engine
  - Search against in-house or evolving public databases
  - Numerous putative ID's

Challenges of current analysis approaches

1. Many separate analysis tools and databases are used, with manual intervention to move data from one stage to the next
2. Little opportunity to visually explore the data
3. You can quantify something that hasn't been identified, and identifications can be ambiguous
4. Discovery experiment needs not as well served as those of a more targeted metabolomics approach, which typically start with knowing what you are after
5. Missing values in the data reduce reliability of statistical tests
6. Not easy to provide a final list of putative compounds for validation and review by biologists
7. **And yours?**

nonlinear DYNAMICS

Progenesis CoMet workflow schematic

CoMet

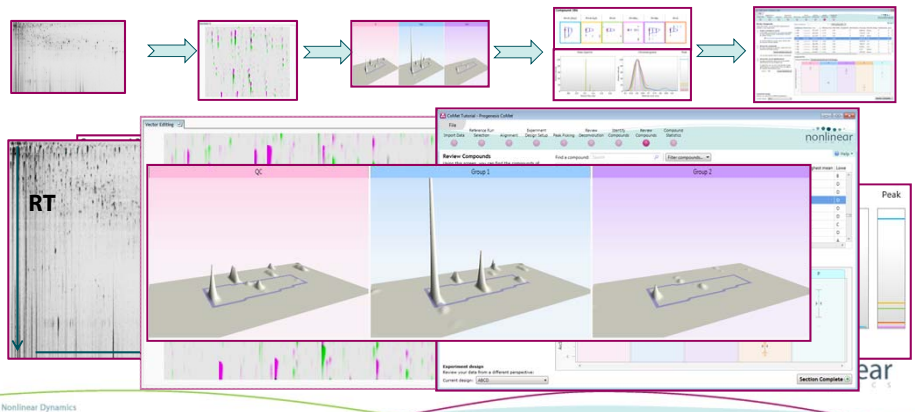
Import Data

Retention time alignment

Co-detect, Quantify and Normalise

Collate ion forms

Identify, Review + Report



nonlinear DYNAMICS

How does it solve the challenges of metabolomics data analysis?

The challenge of missing values

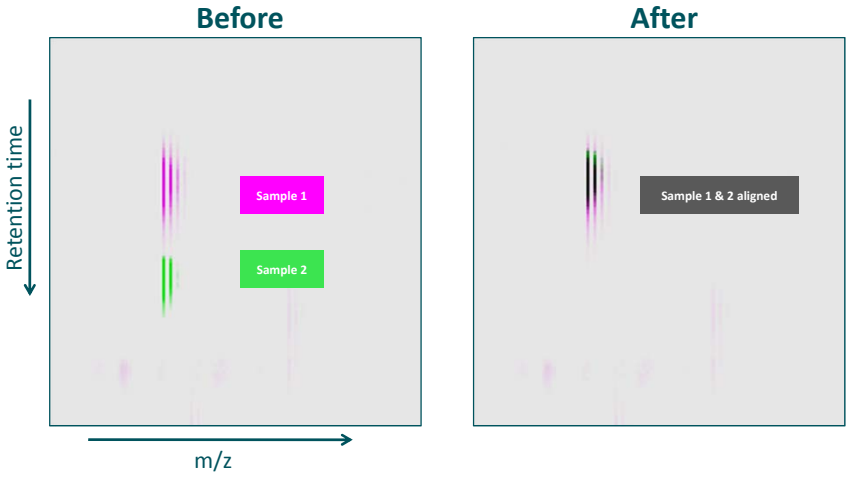
- What are missing values?
  - Any feature that is not matched in every replicate in the experiment

Ref. S...	pi... S /	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot	04... Spot
228	228	196	-	-	208	-	-	-	231	223	-	-	-	185	182	-	-	-	-	-
235	235	-	235	-	-	-	-	163	-	-	-	-	170	189	-	-	-	-	-	-
241	241	202	-	208	226	225	207	231	237	231	159	196	193	-	-	-	-	186	219	
242	242	219	-	-	-	-	-	225	-	232	-	-	201	214	-	201	-	-	221	
263	263	-	263	249	-	-	241	-	281	269	-	-	-	267	270	249	-	224	-	
262	262	262	262	-	-	-	-	-	-	273	-	-	-	-	264	-	-	-	-	
255	255	229	255	-	-	-	-	-	-	272	-	208	-	232	-	-	-	-	226	
256	256	-	256	-	246	-	209	265	-	267	208	-	221	253	-	222	-	-	-	
271	271	232	-	-	-	-	-	272	270	275	-	222	228	261	260	234	178	217	-	

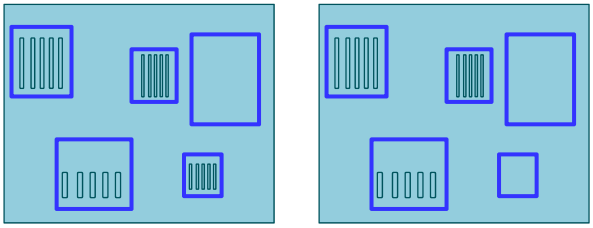
- This is a problem because of the effect on various statistical tools which impact the conclusions drawn
  - Reduced power of the statistical tools

nonlinear DYNAMICS

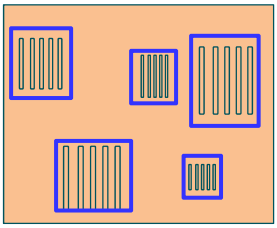
Automatic retention time alignment



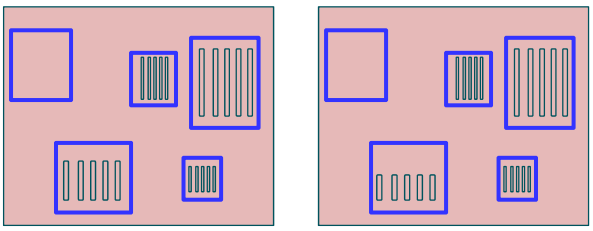
Peak picking and co-detection



Mapping the detection to all runs avoiding missing data

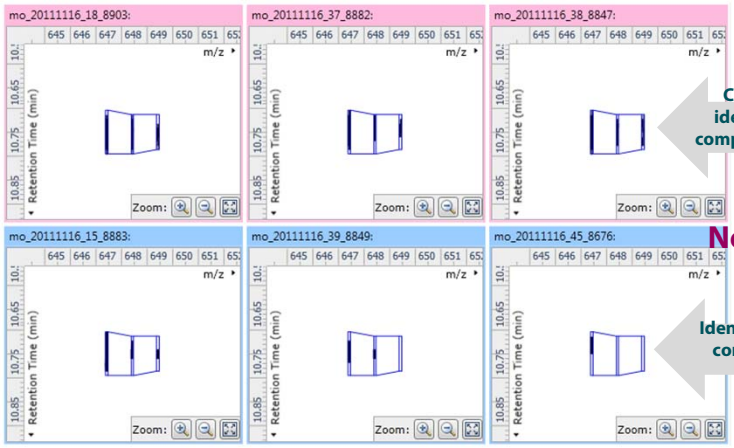


Aggregate co-detection



Co-detection benefits

- Consistent peak picking across all runs
  - **Vital for reliable relative quantification**



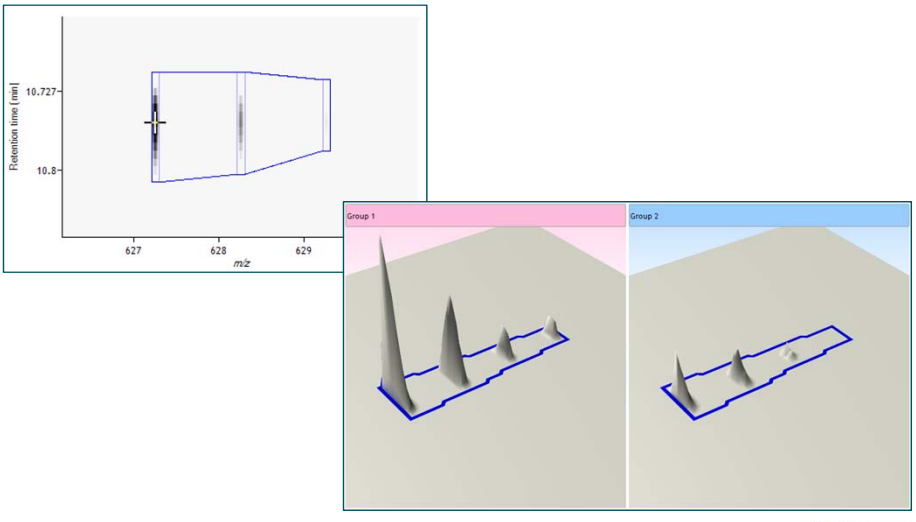
Compound ion identified in one, comparable across all groups

No missing data!

Identification can be confirmed by any other file(s)

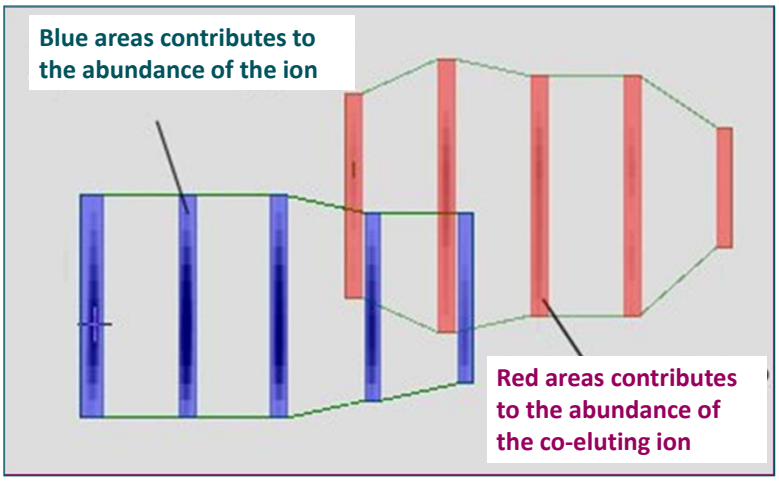
nonlinear DYNAMICS

Compound ion quantification



nonlinear DYNAMICS

Quantify co-eluting ions

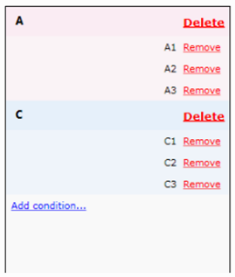


Combining co-eluting compound ions

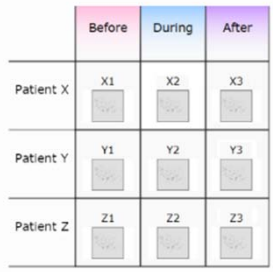
- Deconvolution of co-eluting ion forms
  - Enables the neutral mass to be determined
  - more accurate and reliable compound identifications

Measure expression in multiple groups

- Set up correct experiment design with valid statistics
  - Import experiment design files to **instantly create groups** from hundreds of runs



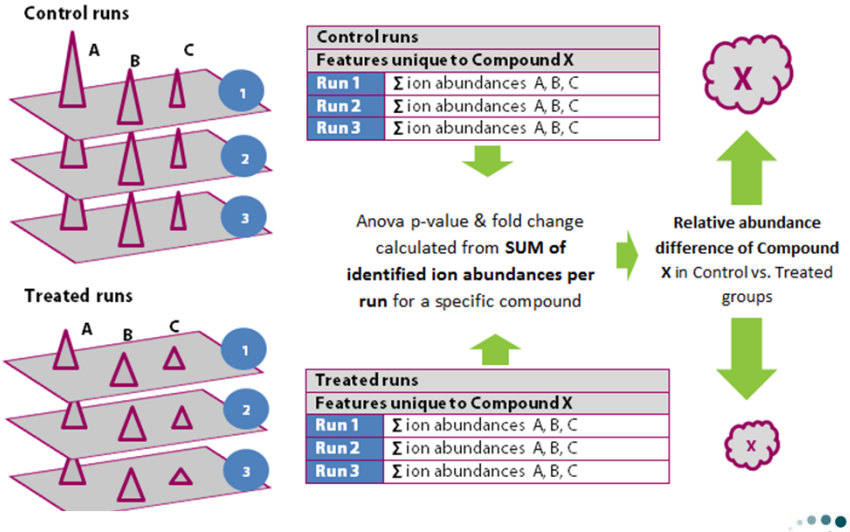
OR



Between-subject Design

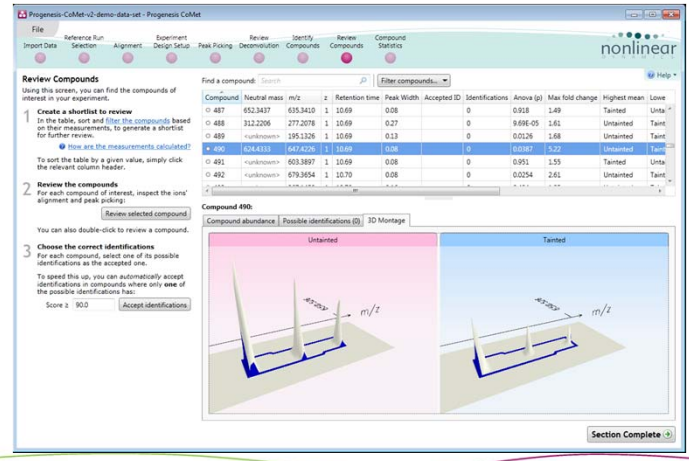
Within-subject Design

Quantifying compound abundances



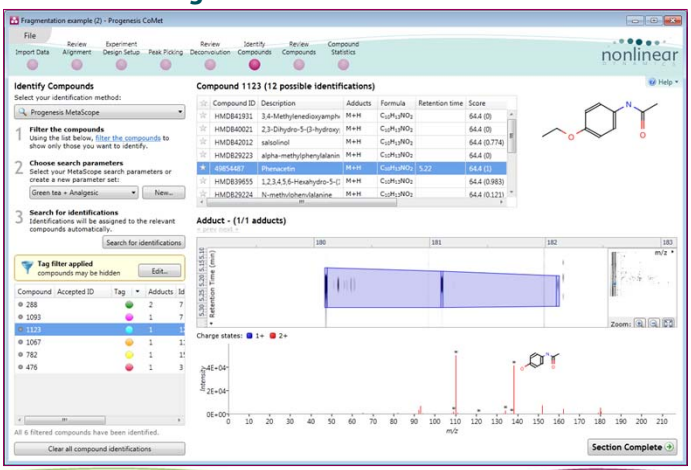
Review compounds

- Quickly find the compounds that are **significantly changing** between any groups
  - e.g. Anova p-value, fold-change, power



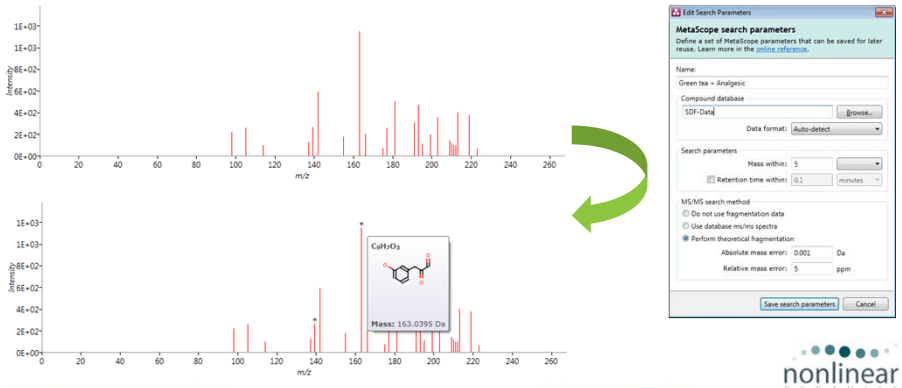
Compound identification

- MetaScope, our search tool, is integrated into the software
  - Search your own data and return compound identifications, **including chemical structures** from SDF databases



Fragmentation support for identification

- **Search SDF databases** using neutral mass, m/z, adduct mass, fragment ion and retention time
- Return putative identifications with **fragment ion matches displayed**

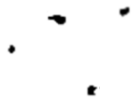


nonlinear DYNAMICS

Progenesis Stats

Easy to use, multivariate statistical analysis

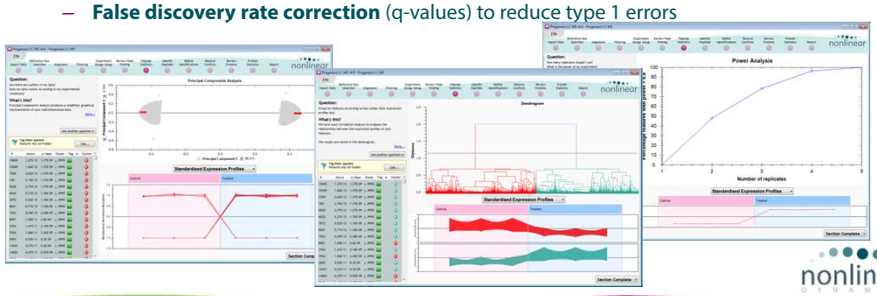
Univariate analysis (e.g. t-test / ANOVA)



Multivariate analysis

Statistical analysis of metabolomics data

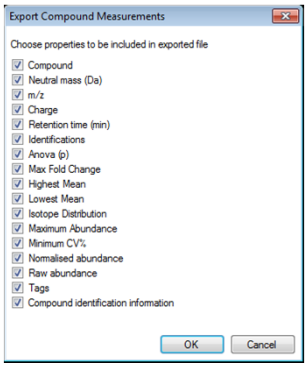
- Find significant Metabolites using statistical tools with a complete matrix of quantitative data (no missing data) incl:
  - **ANOVA** p-values
  - **Principal Components Analysis (PCA)** unsupervised display of how samples group to best describe differences between them, a good QC check!
  - **Correlation analysis** of peptide ions i.e. see which peptides are highly correlated within and across groups
  - **Power analysis per feature** to reduce type II errors
  - **Predictive power** for predicting how many samples are needed to be confident you are not missing something
  - **False discovery rate correction** (q-values) to reduce type 1 errors

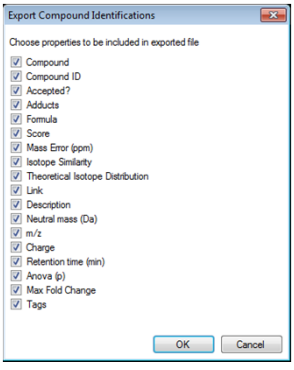


nonlinear DYNAMICS

Put results into biological context

- Comprehensive data export options to complement your bioinformatics-based strategies
  - Compound measurements and identifications
  - Raw abundance of every isotope peak for every adduct

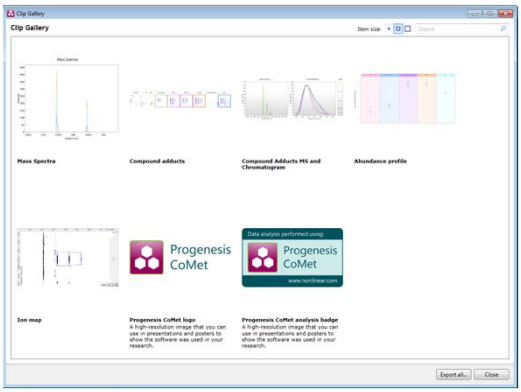




nonlinear DYNAMICS

Capture publication-quality images

- Work through your analysis adding images and tables to a clip gallery
- Select **high quality images for publications, posters and presentations**
  - Saved as high resolution .png files (300 dpi)



nonlinear DYNAMICS

nonlinear DYNAMICS

Progenesis CoMet

Discover the significantly changing compounds in your samples…