



Progenesis CoMet

Concepts and challenges of Metabolomics data analysis.

Mark Bennett – General Manager

24th July, 2013

Copyright Nonlinear Dynamics

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

© Nonlinear Dynamics

Goal of metabolomics discovery?

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

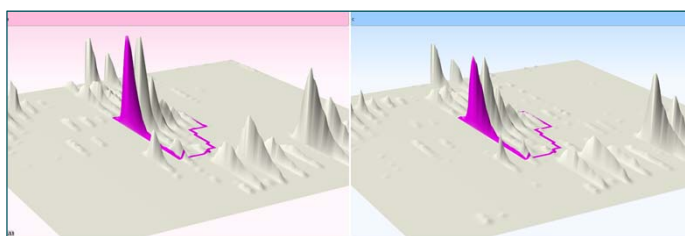
- **Progenesis CoMet** developed with this in mind...
 - Produce a comprehensive **table of detected compounds**, which you can easily share or validate and put them in biological context
 - A complete analysis approach to combine quantification and identification of significantly changing compounds, 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

© 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

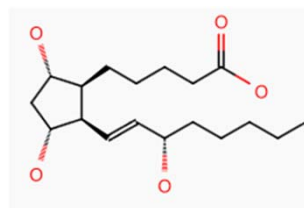


nonlinear
DYNAMICS

© Nonlinear Dynamics

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, \pm RT, \pm MS/MS spectra)
 - No standardised search engine
 - Search against in-house or evolving public databases
 - Numerous putative ID's



nonlinear
DYNAMICS

© Nonlinear Dynamics

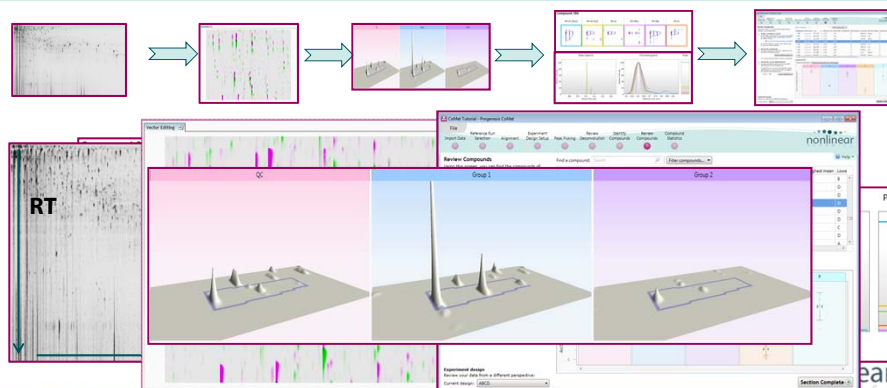
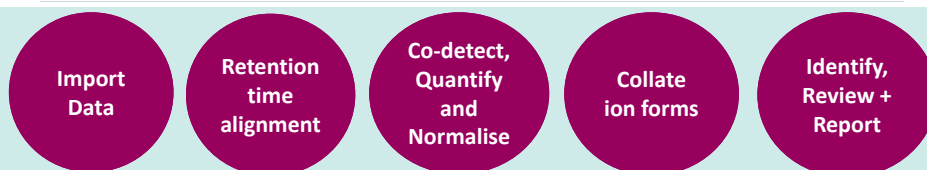
Challenges of current analysis approaches

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

© Nonlinear Dynamics

Progenesis CoMet workflow schematic



© Nonlinear Dynamics



How does it solve the challenges of metabolomics data analysis?



The challenge of missing values

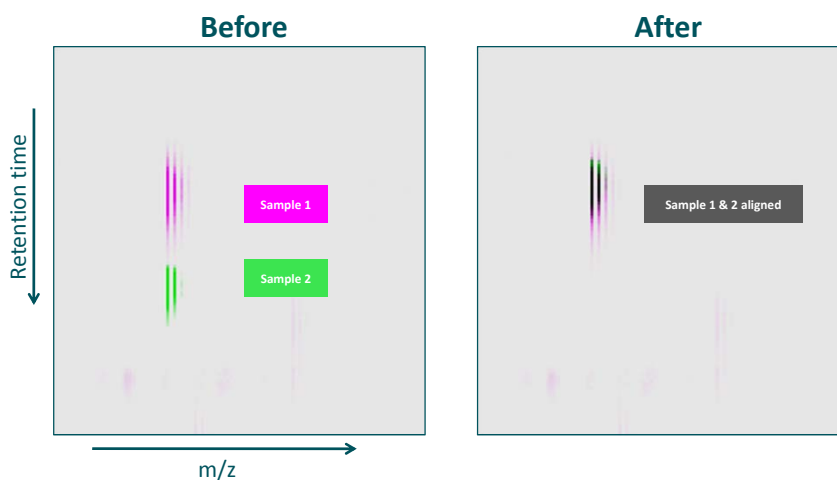
- What are missing values?
 - Any feature that is not matched in every replicate in the experiment

Ref.	pt...	04...	04...	04...	04...	04...	04...	04...	04...	04...	04...	04...	04...	04...	04...	04...	04...
S...	S /	Spot	Spot	Spot	Spot	Spot	Spot	Spot	Spot	Spot	Spot	Spot	Spot	Spot	Spot	Spot	Spot
228	228	196	208	208	208	208	208	208	208	208	208	208	208	208	208	208	208
235	235	-	235	-	-	-	163	-	-	-	-	-	170	189	-	-	-
241	241	202	-	208	226	225	207	231	237	231	159	196	193	-	-	-	186
242	242	219	-	-	-	-	-	225	-	232	-	-	201	214	-	201	-
263	263	-	263	249	-	-	241	-	281	269	-	-	-	267	270	249	-
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



Automatic retention time alignment



nonlinear
DYNAMICS

© Nonlinear Dynamics

Peak picking and co-detection

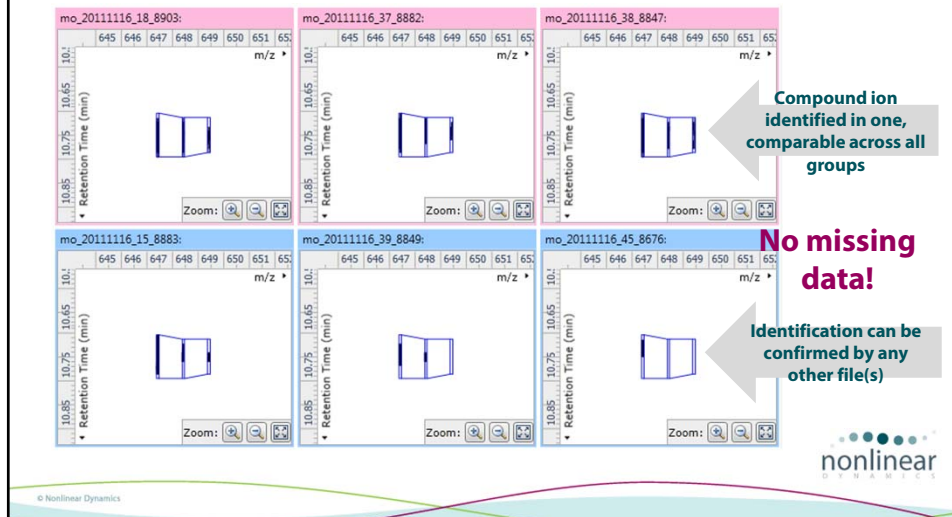


nonlinear
DYNAMICS

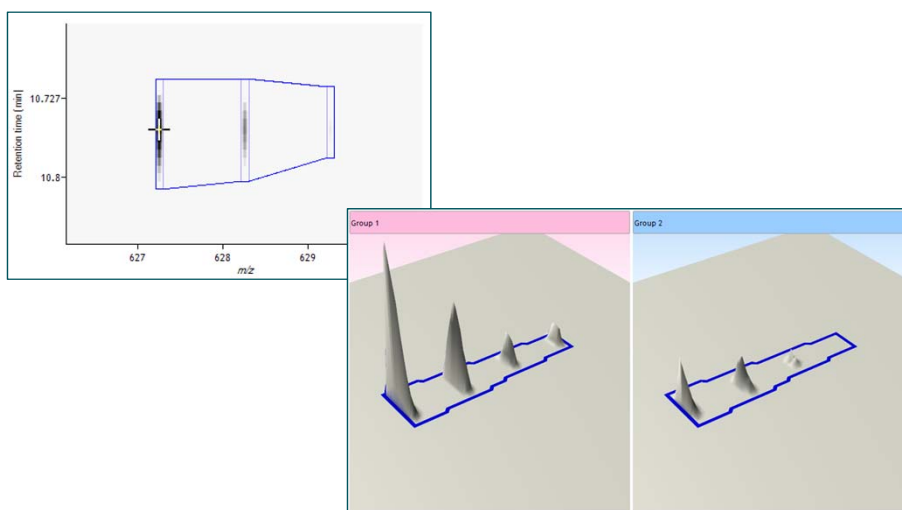
© Nonlinear Dynamics

Co-detection benefits

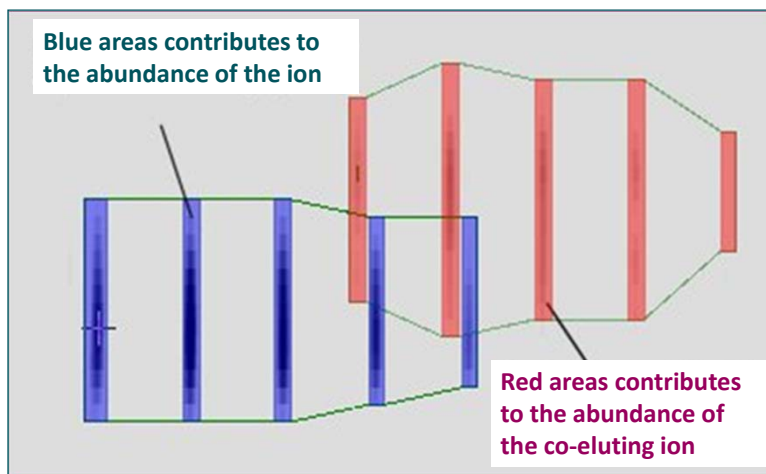
- Consistent peak picking across all runs
 - **Vital for reliable relative quantification**



Compound ion quantification



Quantify co-eluting ions



nonlinear

© Nonlinear Dynamics

Combining co-eluting compound ions

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



nonlinear

© Nonlinear Dynamics

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

A		Delete
A1	Remove	
A2	Remove	
A3	Remove	
C		Delete
C1	Remove	
C2	Remove	
C3	Remove	
Add condition...		

OR

	Before	During	After
Patient X	X1 []	X2 []	X3 []
Patient Y	Y1 []	Y2 []	Y3 []
Patient Z	Z1 []	Z2 []	Z3 []



Between-subject Design

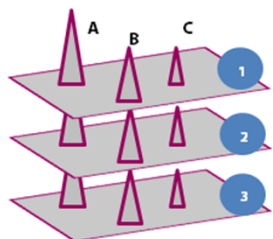


Within-subject Design

© Nonlinear Dynamics

Quantifying compound abundances

Control runs

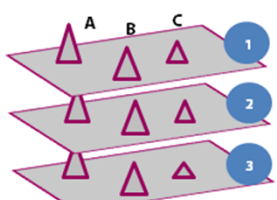


Control runs	
Features unique to Compound X	
Run 1	Σ ion abundances A, B, C
Run 2	Σ ion abundances A, B, C
Run 3	Σ ion abundances A, B, C

Anova p-value & fold change
calculated from **SUM of
identified ion abundances per
run** for a specific compound

Relative abundance
difference of Compound
X in Control vs. Treated
groups

Treated runs



Treated runs	
Features unique to Compound X	
Run 1	Σ ion abundances A, B, C
Run 2	Σ ion abundances A, B, C
Run 3	Σ ion abundances A, B, C

© Nonlinear Dynamics

Review compounds

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

Review Compounds

Using this screen, you can find the compounds of interest in your experiment.

1 **Create a shortlist to review**
In the table, sort and filter the compounds based on their measurements, to generate a shortlist for further review.
• How are the measurements calculated?
To sort the table by a given value, simply click the relevant column header.

2 **Review the compounds**
For each compound of interest, inspect the ions' alignment and peak picking.

3 **Choose the correct identifications**
For each compound, select one of its possible identifications as the accepted one.
To speed this up, you can automatically accept identifications in compounds where only one of the possible identifications has a score ≥ 9.00 .

Compound	Neutral mass	m/z	Retention time	Peak Width	Accepted ID	Identifications	Anova (p)	Max fold change	Highest mean	Lower
487	652.3437	635.3410	1	10.69	0.08	0	0.928	1.49	Tainted	Unta
488	312.2206	277.2078	1	10.69	0.27	0	9.69E-05	1.61	Untainted	Taint
489	unknown	159.1238	1	10.69	0.13	0	0.026	1.68	Untainted	Taint
496	624.4333	607.4206	1	10.69	0.08	0	0.0087	5.22	Untainted	Taint
491	unknown	603.3897	1	10.69	0.08	0	0.951	1.55	Tainted	Unta
492	unknown	679.3654	1	10.70	0.08	0	0.0254	2.61	Untainted	Taint

Compound 496: Compound abundance | Possible identifications (0) | 3D Montage

Score ≥ 9.00

nonlinear
DYNAMICS

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

Identify Compounds

Select your identification method:
Progenesis MetaScope

1 **Filter the compounds**
Using the list below, filter the compounds to show only those you want to identify.

2 **Choose search parameters**
Select your MetaScope search parameters or create a new parameter set:
Green tea - Analgesic

3 **Search for identifications**
Identifications will be assigned to the relevant compounds automatically.

Compound 1123 (12 possible identifications)

Compound ID	Description	Adducts	Formula	Retention time	Score
HMDB01211	2,6-Diaminocaprylamide	M+H	C ₁₂ H ₂₁ N ₂ O ₂	64.4 (0)	64.4 (0)
HMDB04021	2,3-Dihydro-5-(3-hydroxy	M+H	C ₁₂ H ₁₉ N ₂ O ₂	64.4 (0.774)	64.4 (0.774)
HMDB02023	salsolinol	M+H	C ₁₂ H ₁₉ N ₂ O ₂	64.4 (0)	64.4 (0)
HMDB02023	alpha-methylphenylethanol	M+H	C ₁₂ H ₁₉ N ₂ O ₂	64.4 (0)	64.4 (0)
HMDB04497	Phenacetin	M+H	C ₁₂ H ₁₇ N ₂ O ₂	64.4 (1)	64.4 (1)
HMDB09655	2,3,4,5,6-Pentahydro-5-C	M+H	C ₁₂ H ₁₉ N ₂ O ₂	64.4 (0.983)	64.4 (0.983)
HMDB02024	N-methylphenylethanol	M+H	C ₁₂ H ₁₉ N ₂ O ₂	64.4 (0.121)	64.4 (0.121)

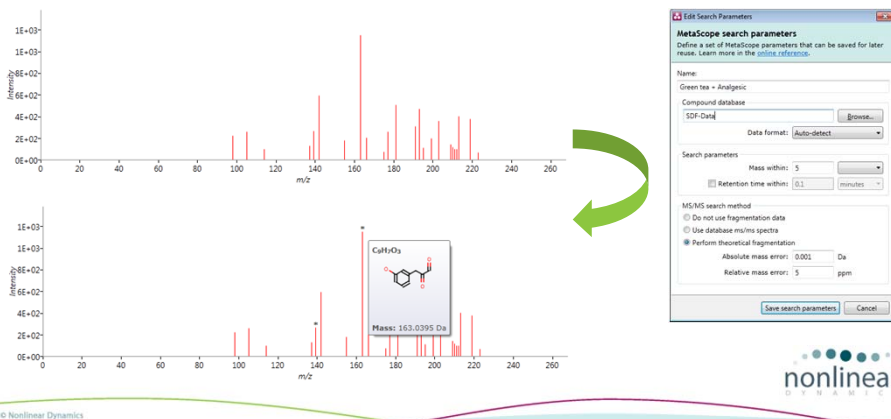
Adduct - (1/1 adducts)

Charge states: 1+ 2+

nonlinear
DYNAMICS

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
D Y N A M I C S

Progenesis Stats

Easy to use, multivariate statistical analysis

Copyright Nonlinear Dynamics

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



nonlinear
DYNAMICS

© Nonlinear Dynamics

Multivariate analysis



nonlinear
DYNAMICS

© Nonlinear Dynamics

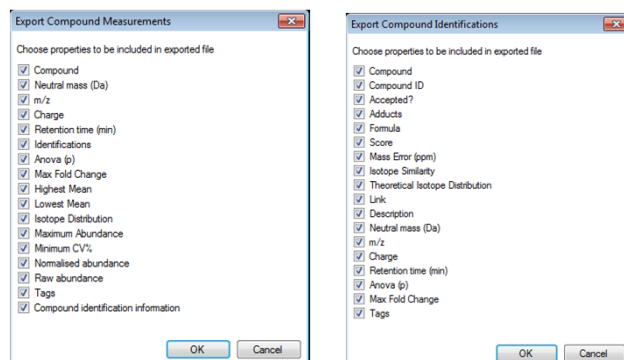
Statistical analysis of metabolomics data

- 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



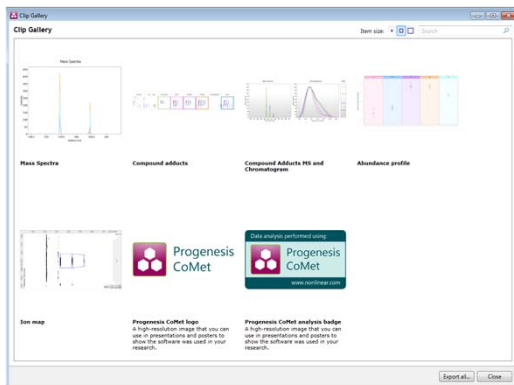
Put results into biological context

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



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

Progenesis CoMet

Discover the significantly changing
compounds in your samples...

Copyright Nonlinear Dynamics