Metabolomic and Proteomics Solutions for Integrated Biology

Christine Miller
Oomics Market Manager
ASMS 2015
Integrating Biological Analysis Using Pathways

- Identifies *why* the pathway is active
- Suggests follow-on experiments
Agilent Instrumentation For Metabolomics

- 7000C GC/QQQ
- 7200B GC/Q-TOF
- 1290 Infinity II UHPLC
- 5977A GC/MS
- Hi-DEF Q-TOF 6500 series
- Q-TOF 6500 series
- QQQ 6400 Series
- TOF 6200 series
- 7100 CE
Most Extensive Software Portfolio for Metabolomics Data Analysis

Find features across many complex data files
Align extraction results from many files
Analyze extracted data for statistically different features
Annotate and identify interesting features
Search and map annotated metabolites to find interesting pathways
Visualize results on pathways
Discovery & Pathway Directed Data Analysis

Discovery
- Separate & Detect
  - LC-TOF/QTOF
- Feature Finding
  - Profinder
- Alignment & Statistics
  - Mass Profiler Professional
- Identify
  - ID Browser
- Pathways
  - Pathway Architect

Pathway Directed
- Separate & Detect
  - LC-TOF/QTOF
- Feature Finding
  - Profinder
- Alignment & Statistics
  - Mass Profiler Professional
- Identify
  - ID Browser
- Pathways
  - Pathway Architect

Agilent Technologies
Lipidomics: The Systematic Study Of The Entire Lipid Profile Of A Cell/Tissue/Organ/Organism

Shotgun lipidomics using QQQ
- Infusion based
- Uses mixed scan functions
- Quantitative using class internal stds
- Class and composition info

Shotgun lipidomics using QTOF
- Infusion based, MS/MS product ion
- Quantitative using class internal stds
- Class, composition and R-group info

Profiling lipidomics
- LC/MS on accurate mass instrument
- MS or MS/MS used for identification
- Class, composition, R-group and isomer

Biological functions include structure, storage, energy and signaling
Agilent Instrumentation For Lipidomics

- 5977A GC/MS
- 7000C GC/QQQ
- 7200B GC/Q-TOF
- Hi-DEF Q-TOF 6500 series
- Q-TOF 6500 series
- 1290 Infinity II UHPLC
- QQQ 6400 Series
- 1260 SFC
- TOF 6200 series
Overview Of Different Chromatographic Separations For Various Lipid Classes

<table>
<thead>
<tr>
<th>Lipid Category</th>
<th>GC/MS</th>
<th>LC/MS</th>
<th>SFC/MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty acids (acyls)</td>
<td>***</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Glycerolipids (triglycerides)</td>
<td>*</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Glycerophospholipids</td>
<td></td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Sphingolipids</td>
<td></td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Sterol lipids</td>
<td>***</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Prenol lipids</td>
<td></td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Saccharolipids</td>
<td></td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Terpenes (plants)</td>
<td>***</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Polyketides</td>
<td>*</td>
<td>**</td>
<td>**</td>
</tr>
</tbody>
</table>

*** is the best choice – chromatography impacts class of lipids resolved and detected!
Agilent Instrumentation For Proteomics

Agilent AssayMAP
Bravo

Agilent Jet Stream source

G1992A Nanospray source

HPLC-Chip

1290 Infinity UHPLC

Q-TOF 6500 series

QQQ 6400 Series

Agilent Technologies
Most Extensive Software Portfolio for Proteomics Data Analysis

- Identify proteins from discovery data-dependent analysis and/or
- Target peptides for proteins of interest
- Analyze data for statistically different features
- Search and map annotated proteins to find interesting pathways
- Visualize results on pathways

- Pathway Analysis
  - Pathway Architect
  - Agilent Bridge DB

- Protein Identification
  - Spectrum Mill

- Target Proteins
  - Skyline
  - Quant

- Mass Profiler Professional

- Differential Analysis

Agilent Technologies
Discovery Proteomics Workflow

Agilent LCMS

Separate & Detect

Identify or Feature Find

Statistics & Visualization

Pathway Analysis

Spectrum Mill Skyline Profinder

Mass Profiler Professional and Pathway Architect

Statistics and visualization

Pathway Architect

Omics data can be analyzed together in the same project
Targeted Proteomics Workflow

Separate & Detect

Target and quantitate

Statistics & Visualization

Pathway Analysis

Agilent LCMS

Q-TOF, QQQ

MassHunter Quant
Spectrum Mill Skyline

Mass Profiler Professional
and Pathway Architect

Statistics and visualization

Pathway Architect

Omics data can be analyzed together in the same project
Integrated Biology Software
GeneSpring Suite

**GX**
- mRNA expression
- Alternative Splicing
- microRNA expression
- Genome-wide association
- Copy Number Variation

**Strand NGS**
- DNA-Seq, RNA-Seq
- ChIP-Seq, Methyl-Seq,
- small RNA-Seq &
- SureSelect QC

**Mass Profiler Professional**
- Proteomics
- Metabolomics

**Pathway Architect**
- Multi-Omic Pathway Analysis
- Canonical and NL Pathways
LC/MS Data Mining

MassHunter Profinder Batch Processor

Single software for untargeted and targeted feature extraction

Designed specifically for the needs of the metabolomics user

Processes many samples!

Minimizes false positive and negative results

Compound Centric: review and manual editing functionalities

Fast processing
Statistics and Pathway Analysis

Mass Profiler Professional (MPP)

Designed for MS data
Performs many types of statistical analysis
• ANOVA, clustering, PCA, class prediction tools, correlation
ID Browser for compound annotation and identification
Export MS/MS target list
Pathway Architect for biological context
What’s Coming for Integrated Biology Workflows?

Peptide-level proteomics capability
• New output file format for Skyline and Spectrum Mill
  - Detailed peptide information exported for proteins
  - Agilent only capability

MassProfiler Professional 14.0
• Supports new peptide-level export
• Enables peptide-level queries, filters and visualization

Profinder 8.0
  - Faster (multi-threaded)
  - GC/MS support
Peptide-Level Export Empowers Deeper Analysis

Filter on peptides by
- Sequence
- PTMs
- Reproducibility

Statistical and correlation analysis at the peptide level

Visualize peptide results
- For proteins across samples
- On pathways

AGILENT TECHNOLOGIES
Profinder 8.0: Scalable Performance and GC/MS Support

GC/MS by MFE:
- Signals corresponding to different ions elute at the same retention time
- Each colored box = a unique co-elution group = a compound

MFE Spectrum
Spectral match 98.54
Library Spectrum

Library matching of MFE component spectrum
Brain Tumor Study: Background

Immunohistochemistry (protein) is the most commonly used method for patient sample evaluation.

Using genomic signatures, glioblastoma multiforme (GBM) tumors have been classified into the following subtypes:

- Classical
- Mesenchymal
- Neural
- Proneural
Brain Tumor Study: Strategy

Research questions

• Can we separate tumors profiled into subgroups using genomic signatures?
• Can we reduce the number of genes in signature without losing classifier performance?

Approach

• Use cancer genome atlas (TCGA) genomic signatures capable of classifying GBM subtypes
• Acquire discovery proteomics data from grade IV GBM tumors without corresponding genomic information as well as control brain samples collected from epilepsy surgeries
• Leverage proteomics results to improve genomics classification
Brain Tumor Study: Workflow

**Genomics**
- Use hierarchical clustering of mRNA profiles to confirm genomic signatures
- Assess data quality and bias using meta data framework analysis
- Convert gene id to protein accession number

**Proteomics**
- Use PCA of discovery proteomics data to confirm tumor vs. control could be distinguished
- Identify a set of differential proteins for the tumor samples
- Use sample-sample correlation to cluster for QC and subtyping in proteomics experiment

**Multi-omics**
- Perform Venn analysis of differential proteins and genes
- Find reduced set of genes for classifying tumor sub-types
Brain Tumor Study: Hierarchical Clustering of Genomic Signatures

Hierarchical clustering of 840 mRNA profiles from 173 GBM tumors

The four GBM subtypes were re-created
Brain Tumor Study: Metadata Analysis of Genomic Signatures

Metadata obtained from TCGA was imported into the project.

Metadata framework analysis of batch and sex was used to assess batch affects or bias.
Brain Tumor Study: Principal Component Analysis of Genomic Signatures

Principal component analysis of confirms subtypes in GBM samples from TCGA
Brain Tumor Study: Quality Control of Discovery Proteomics Data

Sample 5-T

Principle Component Analysis Plot

Sample Clustered Correlation Plot

control
tumor

T – Tumor samples    E – Epilepsy samples
Tumor subgroups demonstrated, using **370 differential proteins**, in both PCA and sample correlation showing the sample heterogeneity.
Brain Tumor Study: Venn Analysis of Genomic and Proteomic Data Set

- 840 Gene Signatures
- 370 Differential Proteins
- 54 Common Proteins and Genes
Brain Tumor Study: Comparison of 54 Genes Identified Using Proteomics Integration to Original Signature

Original Genomic Signature

Genomic-Proteomic Signature

- Neural
- Proneural
- Classical
- Mesenchymal

PC1  PC2  PC3  PC1  PC2  PC3
Brain Tumor Study: Comparison of Original 370 Proteins to 54 Proteins Identified Using Integration

Original 370 Protein List

54 Protein List after Integration
Integrated Biology Workflow Enhancements From Agilent to Accelerate Your Research

Comprehensive -omics workflows for metabolomics, proteomics, and genomics

Multi-variate software tools for integrated biology workflows

Pathway Architect provides data visualization directly on pathways

Peptide-level export and analysis coming in the Fall!

Profinder for GC/MS and enhanced speed coming in the Fall!
Thank you!