METABOLOMICS AS A TOOL FOR CHARACTERIZING THE EXPOSOME:
DETECTION CHALLENGES

Anthony Macherone, Ph.D.
Sr. Scientist, Agilent Technologies, Inc.
Visiting Professor, John Hopkins University SOM
WHAT IS DETECTION?

- Detector-wise: analytical systems are relatively “fixed”
  - There is a great deal of parity across vendors
- Sensitivity specs of model compounds in solvent is not “detection”
- Detection is more how we utilize the system
  - How we implement the methods in real samples
DETECTION IS...

- The ability to differentiate real signals from the din of noise
  - Qualitative and quantitative
  - Reliably and with confidence
- More than just the detector
  - Must consider holistically
Factors that Affect Detection

- Samples and sample preparation
- The analytical platform
- The methodology
  - The exposome requires measurement many of disparate chemotypes
- Data reduction methodologies
  - Qualitative, quantitative, feature extraction, bioinformatics
SAMPLES: COMMON ISSUES

- All these can interfere with detection
  - Poorly collected, stored
  - Improperly identified
  - Unknown presence of EDTA, heparin, etc.
  - Very small sample volumes
    - Difficult to split across platforms
SAMPLE PREPARATION

- Bears strongly on what is detected
  - Polar vs. non-polar solvents
  - Hot or cold extractions
  - Bligh-Dyer

- For metabolomics - minimize
  - Organic solvent protein crash
  - Spin, filter, dry
  - Reconstitute with appropriate solvent
ANALYTICAL PLATFORM

- No one technology is “all encompassing”
- Must leverage technologies
  - Availability
  - Affordability
  - Accountability
  - Ample coverage of chemical space
ESTIMATE OF EXPOSOME SPACE COVERAGE BY COMMON “DETECTORS”
ANALYTICAL PLATFORMS

- LC-technologies
  - ESI +/- with HILIC and RP covers a great deal of exposome space
    - Small polar compounds that respond to ESI
    - Not universal ionization
    - Matrix suppression

- GC-Technologies
  - EI is universal ionization with largest commercial libraries
    - CI & APCI can be very useful
    - Non-polar, volatiles and semi-volatiles
    - May require derivatization

- NMR
  - No real method development required
    - Broad chemotype coverage
    - Highly specific
    - Sensitivity can be an issue
METHODOLOGY CHALLENGES

- Data-driven, knowledge-driven and semi-targeted methods
  - Which and when?
- Accuracy and precision
- Sensitivity and selectivity
- Other challenges:
  - Dynamic range
    - Concentration dependent
    - “in spectrum”
  - Chromatographic resolution
  - Robustness
METHODOLOGY CHALLENGES: DATA-DRIVEN, KNOWLEDGE-DRIVEN AND SEMI-TARGETED METHODS

- **Data-driven**
  - Discovery screening methods, non-hypothesis driven
  - May not detect 70% of the POPs exposome

- **Knowledge-driven**
  - Targeted quantitative methods, hypothesis driven
  - Most sensitive, especially MS/MS

- **Semi-targeted**
  - Combined targeted and screening method
    - Fixed panel of analytes, quantitative
    - Use full spectral information to interrogate data for unknowns
      - Scan, SIM/Scan, MS/MS/scan, triggered MS/MS

Reproduced from Environmental Health Perspectives (http://ehp.niehs.nih.gov/1308015/)
Untargeted, qualitative analyses
  - Typically not validated methods

During method development:
  - Known analytes of fixed quantities in matrix are created
    - Characterize the precision of the method
  - This “performance standard” must be monitored many times during each analytical run
  - Must pass defined metrics
  - Performance standards provide traceable confidence in the precision of the analytical method
METHODOLOGY CHALLENGES: ACCURACY AND PRECISION

- Targeted, quantitative & semi-targeted analyses
  - Generally validated methods

During method development:
- Known analytes of known quantities in matrix must be created
  - Characterize the accuracy and precision of the method
- This “quality control” is bracketed around unknowns
  - Monitored multiple times during each analytical run
- Must pass defined quantifiable quality metrics
- Quality controls provide traceable confidence in the accuracy and precision of the analytical method
METHODOLOGY CHALLENGES: SPECIFICITY

- High mass accuracy and high resolution
  - Offers high specificity and good sensitivity for most drugs, metabolites and dietary compounds
  - Targeted or semi-targeted analysis will be needed to cover much of the POPs in blood exposome space
- Typically, SQ scan < SQ SIM < MS/MS (quad or trap) < HRMS
METHODOLOGY CHALLENGES: HOW MUCH RESOLUTION IS ENOUGH?*

- As much as possible but nothing comes for free
  - Resolving power and resolution are different properties that define instrument and method related performance, respectively.

- Trade offs:
  - Sensitivity reduced as mass resolving power is increased
    - The argument for higher mass resolution may not become persuasive until the molecular weights being measured become significant
  - Spectral acquisition rate.
    - "The problem is (the) required time is too long to allow sufficient spectra to be obtained across a high-resolution (chromatographic) peak for the peak to be properly delineated (sic)."
  - Especially for GC methods
    - Suspect 20K – 50K for most applications

## EU Environmental Specificity Point System

<table>
<thead>
<tr>
<th>Technique</th>
<th># of ions</th>
<th>Specificity Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>MS/MS</td>
<td>1 precursor and 2 product ions</td>
<td>4</td>
</tr>
<tr>
<td>MS/MS</td>
<td>2 precursors each with 1 product ion</td>
<td>5</td>
</tr>
<tr>
<td>HRMS</td>
<td>n</td>
<td>2n</td>
</tr>
</tbody>
</table>

**Key**

- 1 precursor = 1
- 1 product ion = 1.5
METHODOLOGY CHALLENGES: SENSITIVITY

- Really talking about signal-to-noise (S/N)
  - In matrix, typically MS/MS > HRMS
  - Of course, QTOF may overcome

- Many physical ways to improve signal or reduce noise
  - Super-heated drying gas to improve coulombic fission in ESI
  - Multiple capillaries
  - Differentially pumped atmosphere to vacuum chamber
  - Helium quench gas (EI)
  - Off-axis detectors

- Relatively fixed across vendors
  - Comparing like systems
MEASURING MULTIPLE CHEMOTYPES

- Common exposures
  - PM
  - smoke
  - flavinoids
  - minerals
  - lipids
  - amino acids
  - sugars
  - aflatoxins
  - fiber
  - fatty acids
  - APIs
  - excipients
  - volatile solvents
  - POPs
  - etc....

Requires data-driven methodologies
- Generic, broad coverage

Contrary to “traditional” analytical chemistry
DATA REDUCTION

- Bioinformatic algorithms can compare myriad molecules between disease cases / controls
- Pinpoint discriminating exposures
- Availability of commercial libraries
  - Human Metabolome Database currently has more than 41,000 entries
  - Metlin database has tandem mass spectra for more than 64,000 small-molecules
  - NIST14
    - EI mass spectral library: 242,477 unique compounds with chemical structures
    - MS/MS library: 234,284 spectra: 51,216 ion trap spectra of 8,171 compounds and 183,068 collision cell spectra of 7,692 compounds
Targeted quantification
Analysis of a small set of predefined target compounds using a particular technique

Standard compound
Isolation and separation of analytes from matrix
Identification by RT & MS
Calculation of recovery of surrogates
Data analysis calculations from internal standards
Quantification of specific compounds in a sample

Non-targeted discovery
Characterization of broader/unknown compounds through untargeted screening techniques

Sample → QTOF analysis and sample comparisons → Data alignment and analysis → Validation with MS/MS from standards

Compound profile in a sample

LC/MS Triple Quad
Polar or moderately polar compounds

GC/MS Triple Quad
Volatile and semi-volatile compounds

LC/MS QTOF
Polar or moderately polar compounds

GC/MS QTOF
Volatile and semi-volatile compounds

Cell bioassay
Quantifiable cellular responses from mixtures isolated from biological samples

LC/GC ICP/MS
Untargeted organohalogens & complexed metals

ICP/MS
Metals and targeted organic complexes

IC/MS
Anions and cations
Environmental Determinants of Susceptibility

Semi-targeted analysis: Discovery and quantification

Analysis of a small set of predefined target compounds using a particular technique and Characterization of broader/unknown compounds through untargeted screening techniques

Standard compound → Isolation and separation of analytes from matrix → Identification by RT & MS → Calculation of recovery of surrogates → Data analysis calculations from internal standards → Quantification of specific compounds in a sample → Sample → QTOF analysis and sample comparisons → Data alignment and analysis → Validation with MS/MS from standards → Compound profile in a sample

GC/MS Triple Quad
Volatile and semi-volatile compounds

LC/MS QTOF
Polar or moderately polar compounds

GC/MS QTOF
Volatile and semi-volatile compounds

LC/MS Triple Quad
Polar or moderately polar compounds

IC/MS
Anions and cations

LC/GC ICP/MS
Untargeted organohalogens & complexed metals
DATA-DRIVEN LC-QTOF

Reproduced from Environmental Health Perspectives (http://ehp.niehs.nih.gov/1308015/)
Bioinformatics

LC-QTOF
Metabolites, Drugs, Dietary chemicals

Targeted GC-MS/MS
POPs

LC-TOF Methodology described for the analysis of 30K chemical entities from 100 uL sample

Many commercially available methods for analysis of 600 to more than 1000 chemicals in a single sample
Targeted GC-MS/MS
400 - 600 compounds

Semi-targeted GC-QTOF
10s of 1000s

Reproduced from Environmental Health Perspectives (http://ehp.niehs.nih.gov/1308015/)
BLENDED APPLICATIONS

- LC-MS and GC-MS combined: measure > 50K + chemicals
- Broad coverage of the exposome space
  - Add ICP-MS for complexed metals, etc.
- Use bioinformatics tools to combine and align data
FILTERING NORMAL VARIATION FROM THE UNKNOWN SIGNALS OF INTEREST

- Focusing on the blood exposome can efficiently detect exposures from both exogenous and endogenous sources
  - Case / controls from prospective cohorts collected over time
- High mass accuracy and high resolution MS can offer high specificity and good sensitivity for most drugs, metabolites and dietary compounds
- Good method protocols
METHOD BEST PRACTICES: SAMPLES

- Use system performance standards throughout sequence
- Randomize cases and controls
- Block a small, fixed number of case / controls between pooled references and QC’s
  - Use equal numbers of cases / controls (e.g., 10 + 10)
- Use technical replicates: \( N \geq 3 \)
  - Especially for low or poorly distributed chemicals
  - Run sequentially
METHOD BEST PRACTICES: PLATFORM AND METHODOLOGY

- Use multiple techniques for best coverage of the chemical space
- Optimize chromatographic resolution
- Use high res accurate mass for discovery
  - May need targeted and semi-targeted
Detection is:

- A composite of
  - Samples and preparation
  - The analytical platform
  - Methodology
  - Bioinformatics
- Filtering normal variation from the unknown signals of interest
  - Requires measures of performance and confidence
To overcome or mitigate detection challenges:

- Develop robust methods
- Use performance standards / QCs for confidence
- Blend platforms if possible
- Use best method practices
- Sophisticated bioinformatics
  - Use more than one
ACKNOWLEDGEMENTS

- Professor Stephen M. Rappaport. University of California, Berkeley
- Professor Shane A. Snyder. University of Arizona
THANK YOU