Metabolomics Pathway Analysis

Dean P. Jones, Ph.D.
Department of Medicine/Division of Pulmonary, Allergy and Critical Care Medicine
Emory University, Atlanta

No financial COI to disclose

Funding Acknowledgements: NIEHS, NIA, NCI, NHLBI, NIDDK, NIAAA, NIAID, Woodruff Foundation, Emory Dept of Medicine, Georgia Research Alliance

Human disease risk is highly dependent upon cumulative lifelong exposures: Defined by Wild (2005) as the Exposome

10-20% of disease risk

Genome, epigenome

Transcriptome

Proteome

Mechanisms of cell signaling and control

Exposome: Diet, environment, infectious exposures

80-90% of disease risk
Proportion of cancer deaths attributed to various environmental factors (Doll and Peto)

- Infection: 10% (?)
- Geophysical: 3%
- Medical Treatments: 1%
- Consumer Products: <1%
- Pollution: 2%
- Occupational: 4%
- Sexual Behavior: 7%
- Food Additives: <1%
- Unknown: ?%
- Tobacco: 30%
- Alcohol: 3%
- Diet: 35%
Alternate Workflows

**Targeted Metabolomics**
- Select analytic target to test hypothesis
  - Select and test analytic method
  - Perform power calculation; design experiment
  - Conduct experiment
  - Analyze samples and perform statistical analysis

**High-resolution metabolomics**
- Pose scientific question (with or without hypothesis)
  - Select relevant samples
  - Analyze samples by high-resolution MS with advanced data extraction algorithms
  - Use bioinformatic methods and database tools to obtain significant metabolites and pathways
  - Perform MS/MS and co-elution studies to verify metabolites
Nutritional and Environmental Metabolomics

Core nutritional metabolome contains about 2,000 chemicals

Current Metabolomic capabilities: >20,000 “metabolites” in plasma or urine

- Core Nutritional Metabolome
- Non-nutritive Chemicals in Diet
- Microbiome-related Chemicals
- Supplements and Pharmaceuticals
- Commercial Products
- Environmental Chemicals

40 Essential nutrients and about 2000 metabolites formed by enzymes encoded by the genome

Food metabolome

What are the other 18,000 chemicals?

Current Metabolomic capabilities: >20,000 “metabolites” in plasma or urine

- Core Nutritional Metabolome
- Non-nutritive Chemicals in Diet
- Microbiome-related Chemicals
- Supplements and Pharmaceuticals
- Commercial Products
- Environmental Chemicals

40 Essential nutrients and about 2000 metabolites formed by enzymes encoded by the genome

Food metabolome

Plant metabolome >200,000 chemicals

Largely uncharacterized (may be 10-40% of plasma metabolome)

>1000 drugs in use

Environmental metabolome

>10,000 agents used

>80,000 registered with EPA

Jones et al Annu Rev Nutr 2012
**High-resolution metabolomics is becoming practical for routine healthcare**

Resolution approaches that of genomics: >20,000 metabolites

Simple 1-step sample processing can be done anywhere
Relatively rugged instruments can be used in hospitals and larger clinics
10 to 60 min run time; currently $50 to $125/sample, cost decreasing

In conjunction with an online health surveillance and forecasting system, metabolomics analysis could have real-time use in clinical practice

[Image of metabolomics data]

**High-resolution metabolomics data for 174 serum samples**

- **Improved data extraction over most approaches:** 34,768 ions, triplicate analyses
- **Summary for C18:** 19,383 ions  Range of detection over 5 orders of magnitude of intensity
- With triplicate analyses, CV is obtained for each metabolite in each sample:
  - 6,247 had median CV < 10%  Mean intensity of ions with CV <10%; $3.0 x 10^5$

**Intensity**
- Mean: $1.2 x 10^5$
- Median: $2.0 x 10^5$

**CV**
- Median: 14.4%

**Missing values**
- >8000 had <5% missing values (enriched in intermediary metabolites)
- Ions with different missing values appear to reflect variable dietary and environmental agents

[Histograms of intensity, CV, and missing values]

Sample analysis (C18) shows that there is consistent LCMS system response during weeklong analysis period.

Triplicate analyses are reproducible

Random high response probably represents a residual problem with sample processing or inconsistencies in complex system (sample, autosampler and injector components, multiple valves, electrospray inhomogeneities or electromagnetic properties of ion transfer tube)

Averages of replicates includes random variation of individual high values but otherwise shows that individuals have relatively consistent total signal.

It is not clear whether individuals have differences in total signal due to amount of total metabolites—we concluded in our earlier NMR studies of SAA insufficiency that this occurred, perhaps due to differences in amount of albumin

An alternative possibility is that there are specific chemicals in some individuals that have global effects on ionization. This could be NaCl content, phosphate, sulfate, total lipid or other high-abundance chemical. Expectation is that it would be principally impacting the initial (salt) washthrough
Key components of pathway analysis

Statistical testing: FDR

Metabolite-metabolite Correlation Analyses

Online Databases/Resources

Cross-platform Studies

**Manhattan plot**: Y axis represents the negative log\(_{10}\) of p-value (higher is better) and the x-axis represents the measured m/z

![Manhattan plot diagram](image)
Metabolite correlations are very useful to understand redundancies of chemical detection and network associations of metabolism.

Detected m/z features matching half of known human intermediary metabolites (KEGG) are shown in black; most human metabolic pathways are represented.
Pathway Analysis of 400 matched m/z significantly different between healthy controls and patients

Pathway Analysis of 35 features contributing to PCR correlation to disease score
Correlations of amino acids with others that share common transport systems demonstrates principle that related metabolites tend to associate with each other when compared across a set of samples.

Analyses of clusters of metabolites reveals clusters of lipids, metabolites related by transport systems and common metabolic enzymes, etc.

How do we deal with massive amount of complex data?

Metabolome-wide association study (MWAS) of BMI
Controlled for age, sex and race/ethnicity
Hierarchical cluster analysis of subjects according to most significant metabolites that differ according to BMI

Two major clusters of individuals

Major clusters of metabolites

Two include High levels of blood lipids

BMI; yellow = high

Pathway enrichment analysis using Metacore

1. Acetylcholine biosynthesis and metabolism
2. Phospholipid metabolism p.2
3. Phospholipid metabolism p.3
4. Ganglioside Metabolism p1
5. Development_Peak of nicotinamide in G-CSF-induced granulopoiesis
6. Vitamin K metabolism
7. N-Acylethanolamines, HSRL5-transacylation pathway
8. N-Acylethanolamines, N-Acyltransferase pathway
9. Vitamin B3 (nicotinamide) metabolism
10. Neurophysiological process_Circadian rhythm
Acetylcholine biosynthesis and metabolism

C00350; m/z 724.4745
C00157; m/z 832.6778
C00307; m/z 552.1

Metscape: pathway mapping of discriminatory metabolites in disease
Urea cycle and metabolism of arginine, proline, glutamate, aspartate, and asparagine

m/z=196.0596; M+H adduct; KEGG: C05932
m/z=219.0966; M+H adduct; KEGG: C003740
m/z=204.1223; M+H adduct; KEGG: C02571
m/z=144.1013; M+H-H2O adduct; KEGG: C00049
m/z=134.0441; M+H adduct; KEGG: C00402
m/z=147.0757; M+H adduct; KEGG: C00064
MetPA provides plot of significance and impact of metabolic pathways to separation of biological classes of samples

In this MetPA analysis, 199 matches to KEGG human compounds from 335 features significant by FDR (q=0.01) shows calculated impact of sphingolipid, vitamin B6 and amino acid metabolism.

Mummichog combines metabolite prediction and network analysis in one step

Shuzhao Li et al, 2013 PLoS Computational Biology
Activity network predicted by *Mummichog* in innate immune activation

![Network Diagram](image1)

Shuzhao Li et al, PLoS Computational Biology 2013

Development of Deconvolution MS/MS for Identification of Low-Abundance Ions

![Graphical Representation](image2)

Uppal et al BMC Bioinformatics 2013
Development of Deconvolution MS/MS for Identification of Low-Abundance Ions

Integrated omics for pathway analysis

Genome x Metabolome
Transcriptome x Metabolome
Proteome x Metabolome
Gene-metabolome (G x M) associations show that metabolites vary in association with disease risk variations in SNPs.

Integrated omics can revolutionize mechanistic research.

One experiment can reveal individual associations of >10,000 metabolites with >10,000 transcripts.
Data can be mapped to network structure to give global picture of system response

Top 500 hub genes with significant metabolite associations in toxicity study

2 Major toxic response networks

Positive association
Negative association

Metscape uses knowledge-based approach to map pathway interactions

Proteome x Metabolome

P x M
Toward a Surveillance & Forecasting System for Personalized Medicine

Have an affordable system using metabolomics to

forecast risk of disease,
timing of disease onset and
intensity of impact

and use this system to

Improve disease prevention,
classification
and treatment