Systems Biology of Human Metabolism and Gut Microbiome – SysMed 2020

Metabolomics

From data to insights

Dr. Abdellah TEBANI

Stockholm - SicLifeLab - 18th February 2020

Division of Systems Biology - Department of Protein Science
Science for Life Laboratory - KTH - Stockholm, Sweden

KTH VETENSKAP OCH KONST

THE HUMAN PROTEIN ATLAS

SciLifeLab

Learning Objectives

You will learn about

- Metabolomics / Targeted and Untargeted
- Analytical technologies
- Metabolomics applications & Limits
- Main processing pipelines
- Tips, pitfalls and traps

Outline

What is metabolomics?

Experimental Design

Data Generation / Analytical technologies

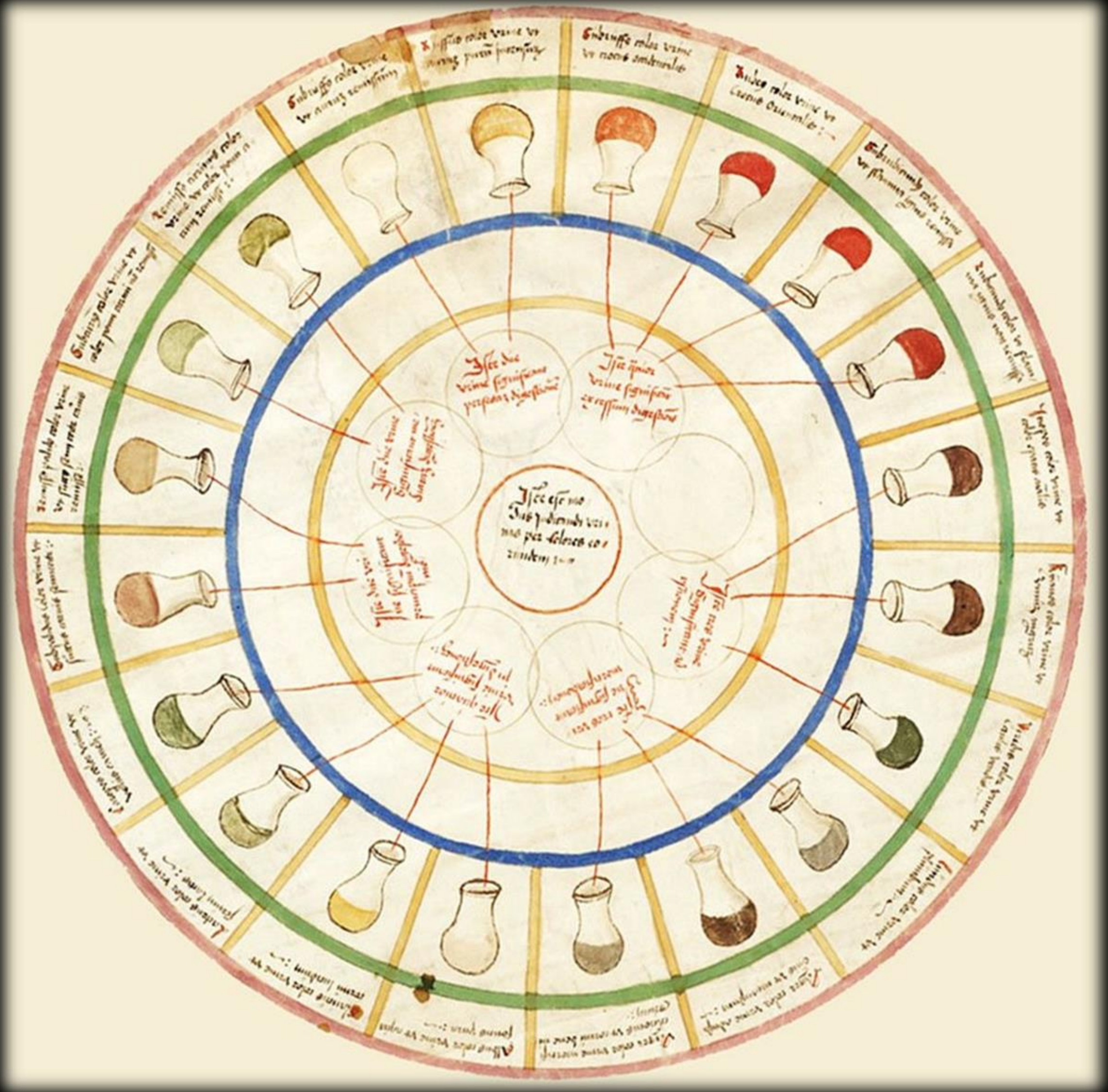
Applications and Limits

Preprocessing

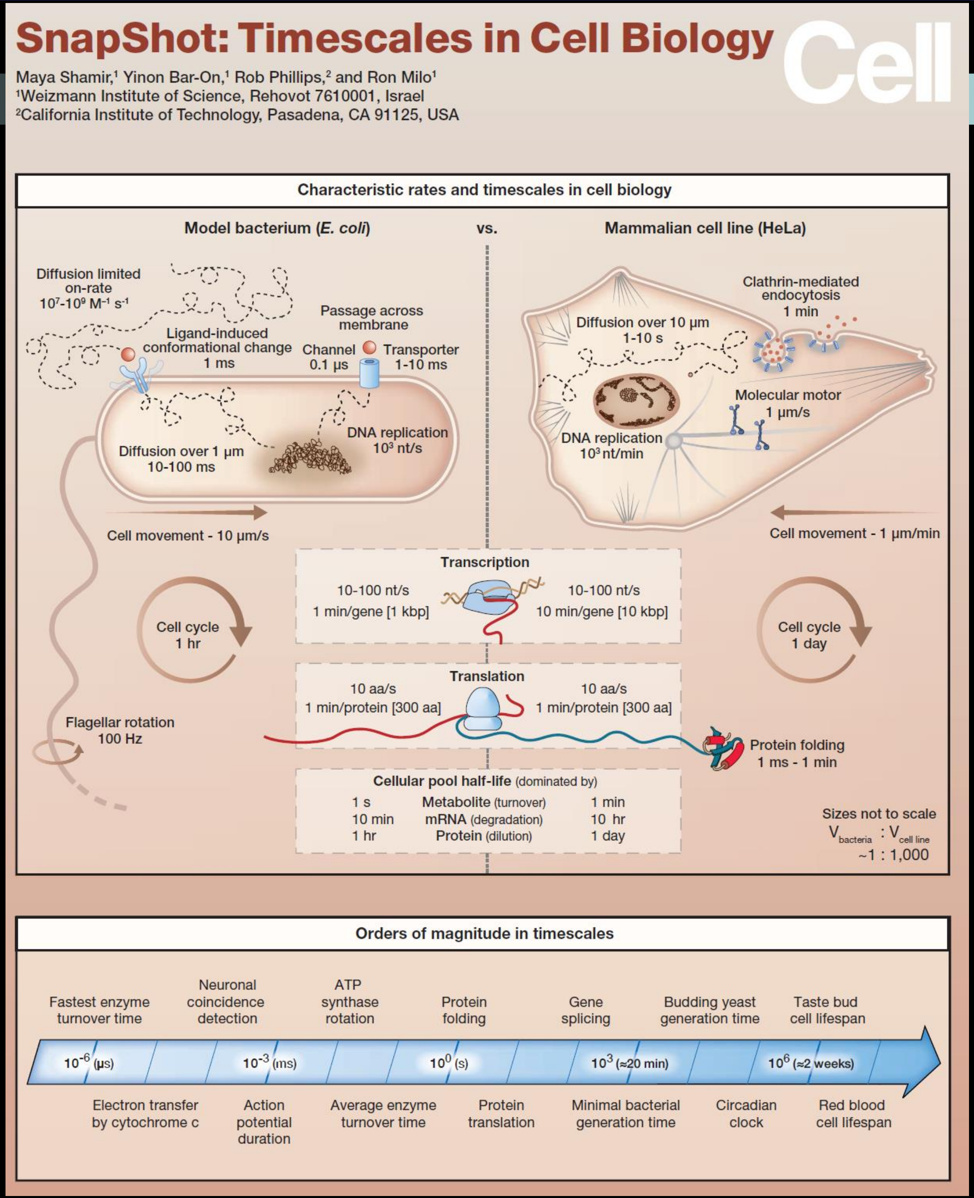
Data cleaning

Key steps and best practice

The urine wheel



Time scales in biology

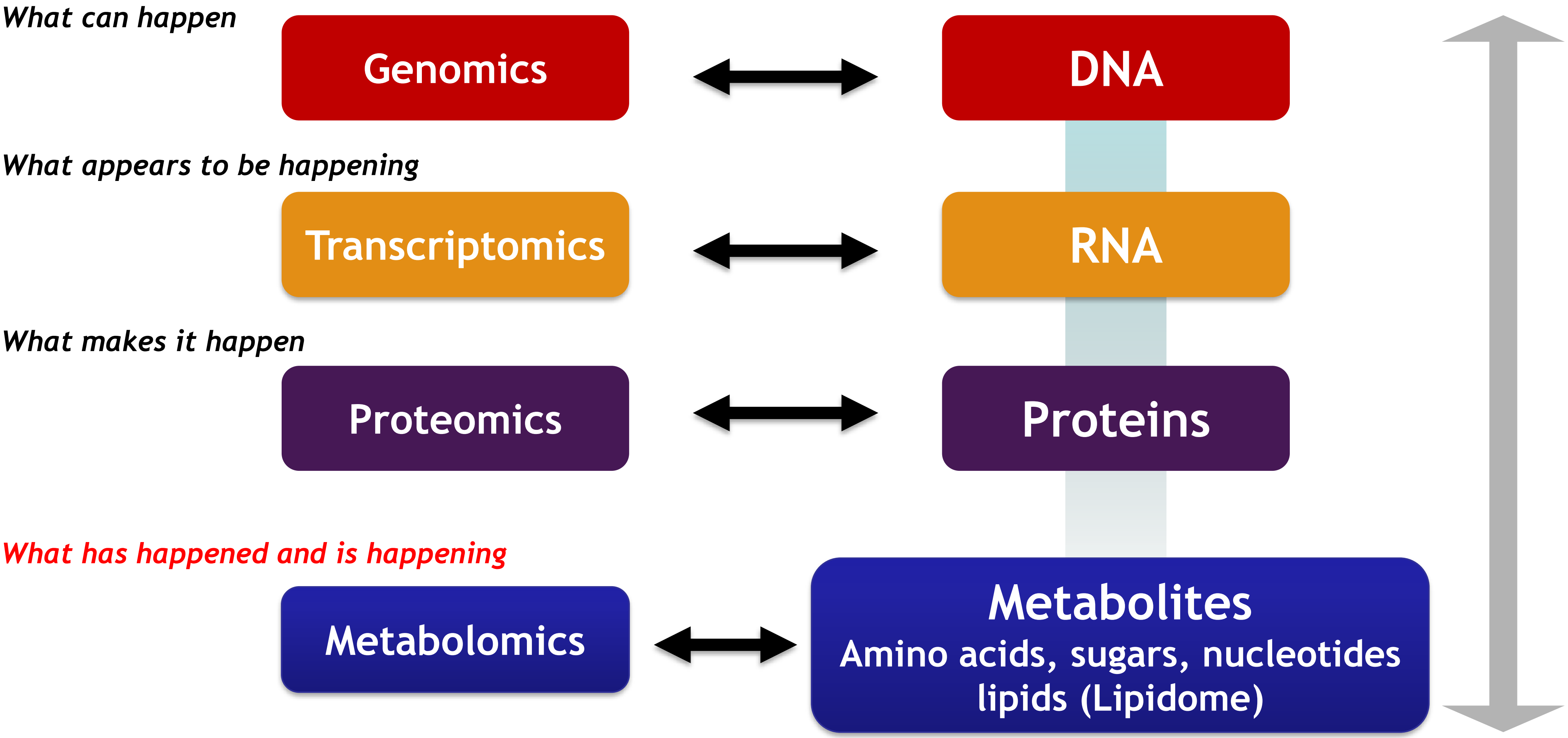


SnapShot: Timescales in Cell Biology

Maya Shamir,[1] Yinon Bar-On,[1] Rob Phillips,[2] and Ron Milo[1]
[1]Weizmann Institute of Science, Rehovot 7610001, Israel
[2]California Institute of Technology, Pasadena, CA 91125, USA

Cell
Characteristic rates and timescales in cell biology
Model bacterium (E. coli)
vs.
Mammalian cell line (HeLa)
Diffusion limited on-rate 10^7-10^9 M^{-1} s^{-1}
Ligand-induced conformational change 1 ms
Passage across membrane
Channel 0.1 µs
Transporter 1-10 ms
DNA replication 10^3 nt/s
Diffusion over 1 µm 10-100 ms
Cell movement - 10 µm/s
Diffusion over 10 µm 1-10 s
Clathrin-mediated endocytosis 1 min
Molecular motor 1 µm/s
DNA replication 10^3 nt/min
Cell movement - 1 µm/min
Transcription
10-100 nt/s
1 min/gene [1 kbp]
10-100 nt/s
10 min/gene [10 kbp]
Cell cycle 1 hr
Cell cycle 1 day
Translation
10 aa/s
1 min/protein [300 aa]
10 aa/s
1 min/protein [300 aa]
Flagellar rotation 100 Hz
Protein folding 1 ms - 1 min
Cellular pool half-life (dominated by)
1 s Metabolite (turnover) 1 min
10 min mRNA (degradation) 10 hr
1 hr Protein (dilution) 1 day
Sizes not to scale
$V_{bacteria}$: $V_{cell\ line}$
~1 : 1,000
Orders of magnitude in timescales
Fastest enzyme turnover time
Neuronal coincidence detection
ATP synthase rotation
Protein folding
Gene splicing
Budding yeast generation time
Taste bud cell lifespan
10^{-6} (µs)
10^{-3} (ms)
10^0 (s)
10^3 (≈20 min)
10^6 (≈2 weeks)
Electron transfer by cytochrome c
Action potential duration
Average enzyme turnover time
Protein translation
Minimal bacterial generation time
Circadian clock
Red blood cell lifespan

Metabolomics



What can happen

Genomics ↔ DNA

What appears to be happening

Transcriptomics ↔ RNA

What makes it happen

Proteomics ↔ Proteins

What has happened and is happening

Metabolomics ↔ Metabolites
Amino acids, sugars, nucleotides
lipids (Lipidome)

What is a metabolite?

Metabolites:

Small molecules (<1500 Da)
Ultimate support of the biological information
Includes human & microbial products
Endogenous metabolites: produced by the host organism
Exogenous metabolites: not produced by the host organism

Metabolome:

refers to the complete set of metabolites in a biological sample

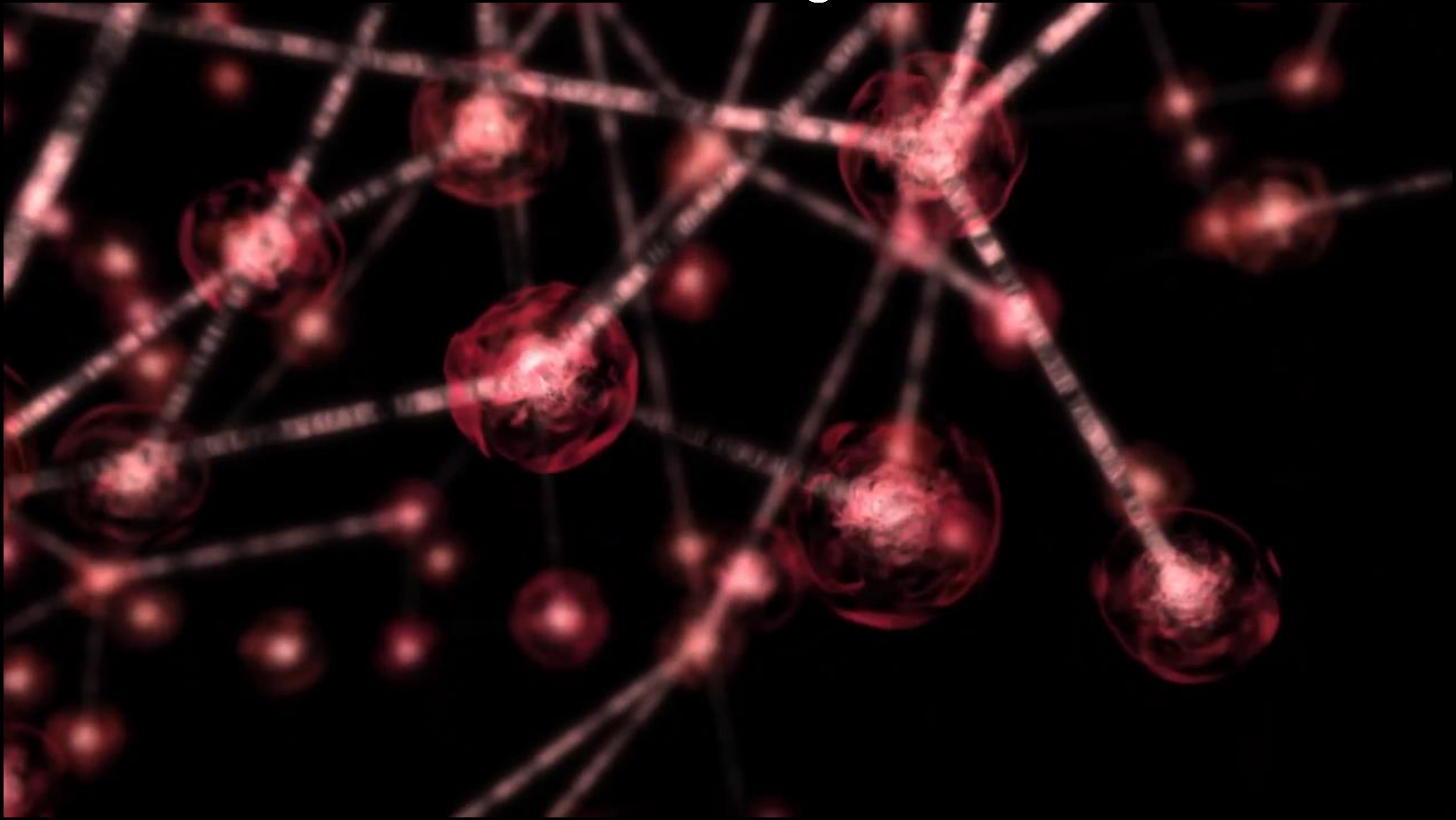
MetaboLomics / MetaboNomics:

➔ **Metabolic Profiling / Metabotyping**

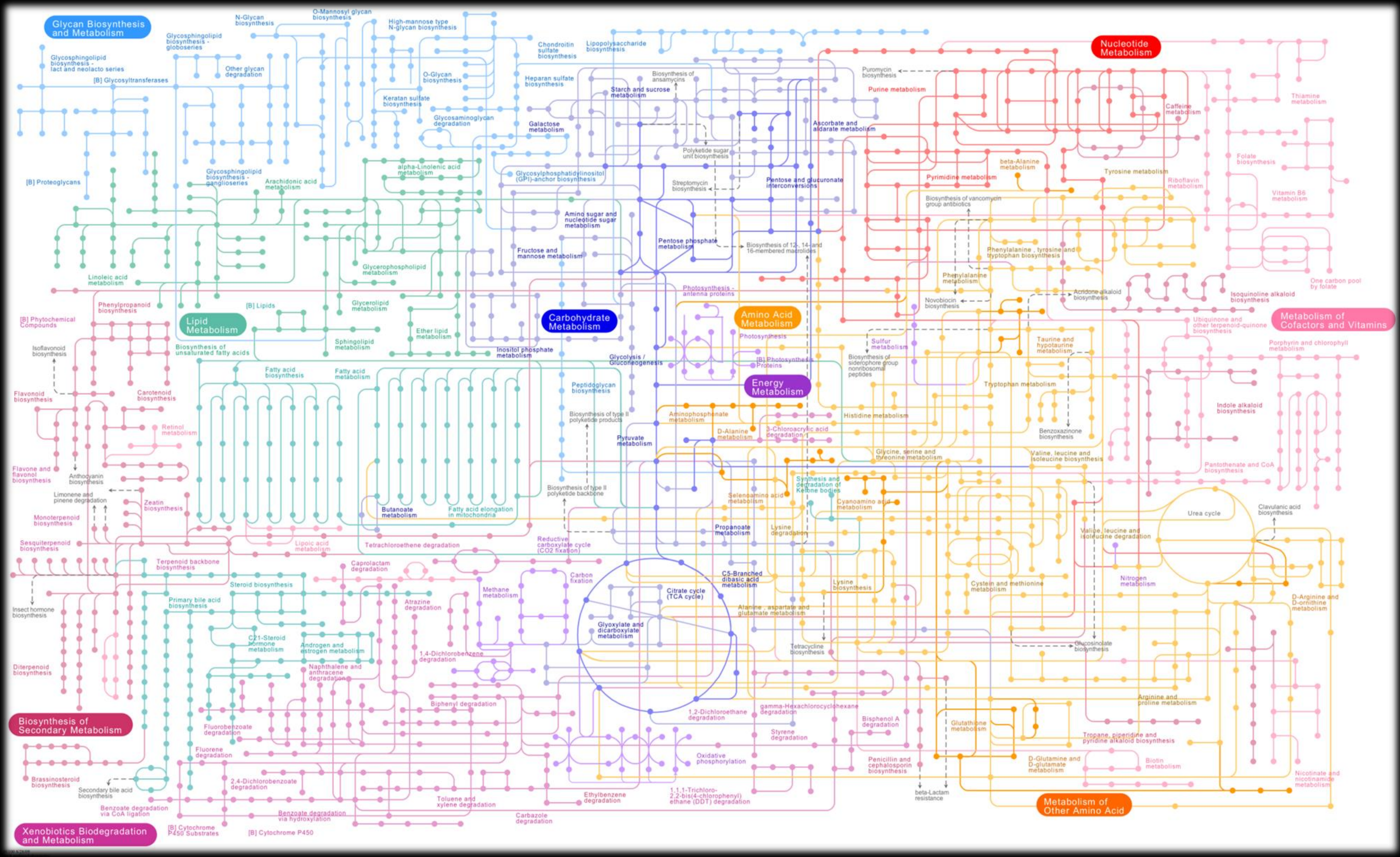
"the quantitative measurement of the metabolic ***responses*** of ***complex systems*** to a pathophysiological ***stimulus*** or genetic modification".

(Nicholson, J. K., et al 1999, Xenobiotica, 29, 1181-89.)

Metabolisme is an integrated network

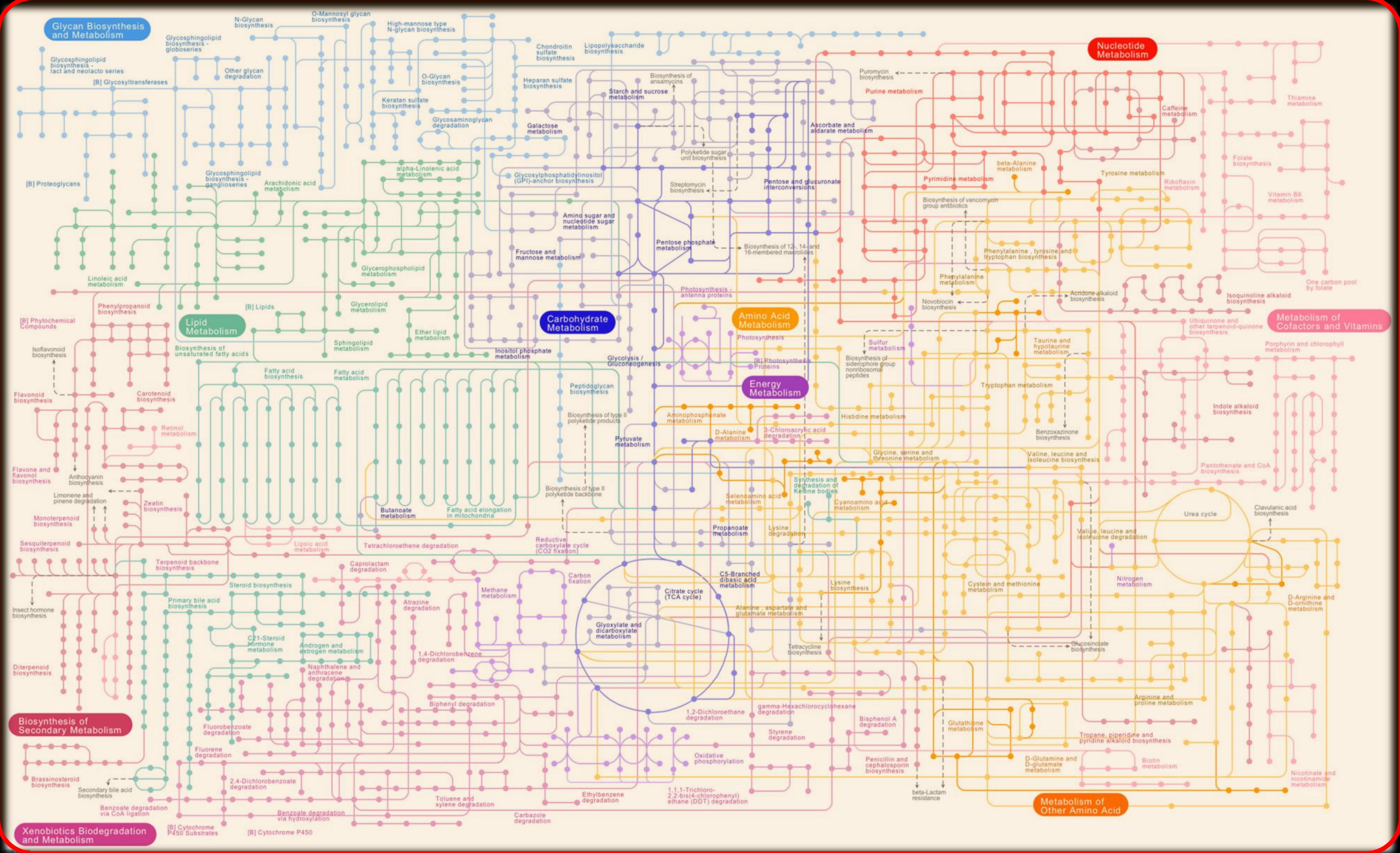


Metabolomics and Metabolism



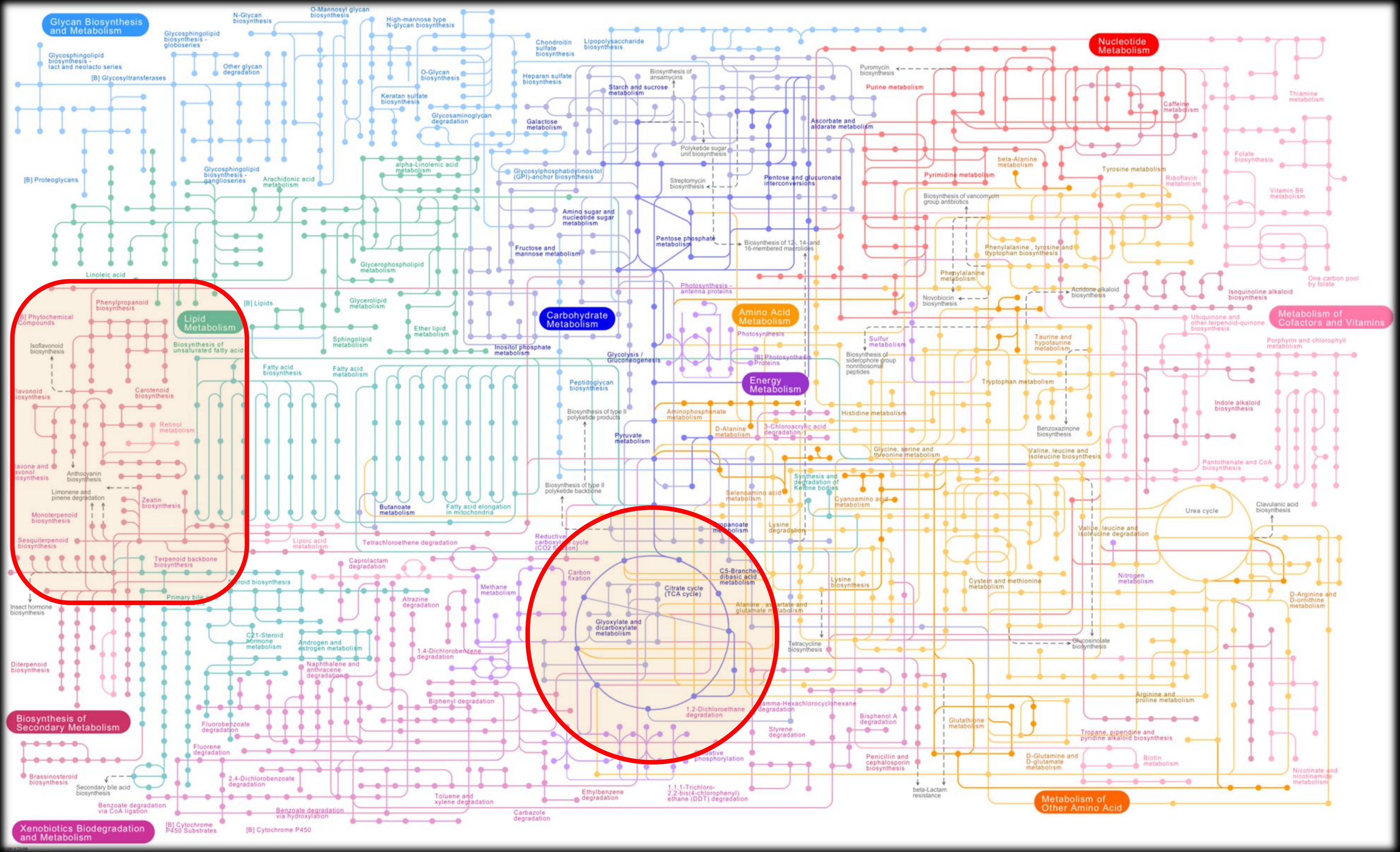
Glycan Biosynthesis and Metabolism
N-Glycan biosynthesis
O-Mannosyl glycan biosynthesis
High-mannose type N-glycan biosynthesis
Glycosphingolipid biosynthesis - globoseries
Glycosphingolipid biosynthesis - lact and neolacto series
Other glycan degradation
[B] Glycosyltransferases
O-Glycan biosynthesis
Keratan sulfate biosynthesis
Glycosaminoglycan degradation
Chondroitin sulfate biosynthesis
Lipopolysaccharide biosynthesis
Heparan sulfate biosynthesis
Galactose metabolism
Biosynthesis of ansamycins
Starch and sucrose metabolism
Puromycin biosynthesis
Purine metabolism
Nucleotide Metabolism
Thiamine metabolism
Caffeine metabolism
Ascorbate and aldarate metabolism
Polyketide sugar unit biosynthesis
beta-Alanine metabolism
Tyrosine metabolism
Folate biosynthesis
Riboflavin metabolism
Vitamin B6 metabolism
Pyrimidine metabolism
Pentose and glucuronate interconversions
Streptomycin biosynthesis
Biosynthesis of vancomycin group antibiotics
[B] Proteoglycans
Glycosphingolipid biosynthesis - ganglioseries
Arachidonic acid metabolism
alpha-Linolenic acid metabolism
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
Amino sugar and nucleotide sugar metabolism
Fructose and mannose metabolism
Pentose phosphate metabolism
Biosynthesis of 12-, 14- and 16-membered macrolides
Phenylalanine, tyrosine and tryptophan biosynthesis
Glycerophospholipid metabolism
Linoleic acid metabolism
Phenylalanine metabolism
One carbon pool by folate
Photosynthesis - antenna proteins
Acridone alkaloid biosynthesis
Isoquinoline alkaloid biosynthesis
Novobiocin biosynthesis
Phenylpropanoid biosynthesis
[B] Lipids
Glycerolipid metabolism
[B] Phytochemical compounds
Lipid Metabolism
Ether lipid metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Ubiquinone and other terpenoid-quinone biosynthesis
Metabolism of Cofactors and Vitamins
Photosynthesis
Sulfur metabolism
Taurine and hypotaurine metabolism
Porphyrin and chlorophyll metabolism
Isoflavonoid biosynthesis
Biosynthesis of unsaturated fatty acids
Sphingolipid metabolism
Inositol phosphate metabolism
Glycolysis / Gluconeogenesis
[B] Photosynthesis Proteins
Biosynthesis of siderophore group nonribosomal peptides
Fatty acid biosynthesis
Fatty acid metabolism
Peptidoglycan biosynthesis
Energy Metabolism
Tryptophan metabolism
Flavonoid biosynthesis
Carotenoid biosynthesis
Biosynthesis of type II polyketide products
Aminophosphonate metabolism
Histidine metabolism
Indole alkaloid biosynthesis
Retinol metabolism
D-Alanine metabolism
3-Chloroacrylic acid degradation
Benzoxazinone biosynthesis
Pyruvate metabolism
Glycine, serine and threonine metabolism
Valine, leucine and isoleucine biosynthesis
Flavone and flavonol biosynthesis
Anthocyanin biosynthesis
Synthesis and degradation of Ketone bodies
Pantothenate and CoA biosynthesis
Biosynthesis of type II polyketide backbone
Limonene and pinene degradation
Selenoamino acid metabolism
Cyanoamino acid metabolism
Zeatin biosynthesis
Butanoate metabolism
Fatty acid elongation in mitochondria
Urea cycle
Clavulanic acid biosynthesis
Monoterpenoid biosynthesis
Propanoate metabolism
Lysine degradation
Valine, leucine and isoleucine degradation
Sesquiterpenoid biosynthesis
Lipoic acid metabolism
Tetrachloroethene degradation
Reductive carboxylate cycle (CO2 fixation)
Terpenoid backbone biosynthesis
Caprolactam degradation
C5-Branched dibasic acid metabolism
Carbon fixation
Lysine biosynthesis
Cystein and methionine metabolism
Nitrogen metabolism
Steroid biosynthesis
Methane metabolism
Citrate cycle (TCA cycle)
Insect hormone biosynthesis
Primary bile acid biosynthesis
Atrazine degradation
Alanine, aspartate and glutamate metabolism
D-Arginine and D-ornithine metabolism
Glyoxylate and dicarboxylate metabolism
C21-Steroid hormone metabolism
Androgen and estrogen metabolism
Tetracycline biosynthesis
Glucosinolate biosynthesis
1,4-Dichlorobenzene degradation
Diterpenoid biosynthesis
Naphthalene and anthracene degradation
Arginine and proline metabolism
Biphenyl degradation
gamma-Hexachlorocyclohexane degradation
1,2-Dichloroethane degradation
Bisphenol A degradation
Glutathione metabolism
Biosynthesis of Secondary Metabolism
Fluorobenzoate degradation
Styrene degradation
Tropane, piperidine and pyridine alkaloid biosynthesis
Fluorene degradation
Oxidative phosphorylation
Penicillin and cephalosporin biosynthesis
D-Glutamine and D-glutamate metabolism
Biotin metabolism
Brassinosteroid biosynthesis
2,4-Dichlorobenzoate degradation
Nicotinate and nicotinamide metabolism
Secondary bile acid biosynthesis
1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT) degradation
beta-Lactam resistance
Benzoate degradation via CoA ligation
Toluene and xylene degradation
Ethylbenzene degradation
Metabolism of Other Amino Acid
Benzoate degradation via hydroxylation
Carbazole degradation
Xenobiotics Biodegradation and Metabolism
[B] Cytochrome P450 Substrates
[B] Cytochrome P450

Metabolomics (Untargeted)

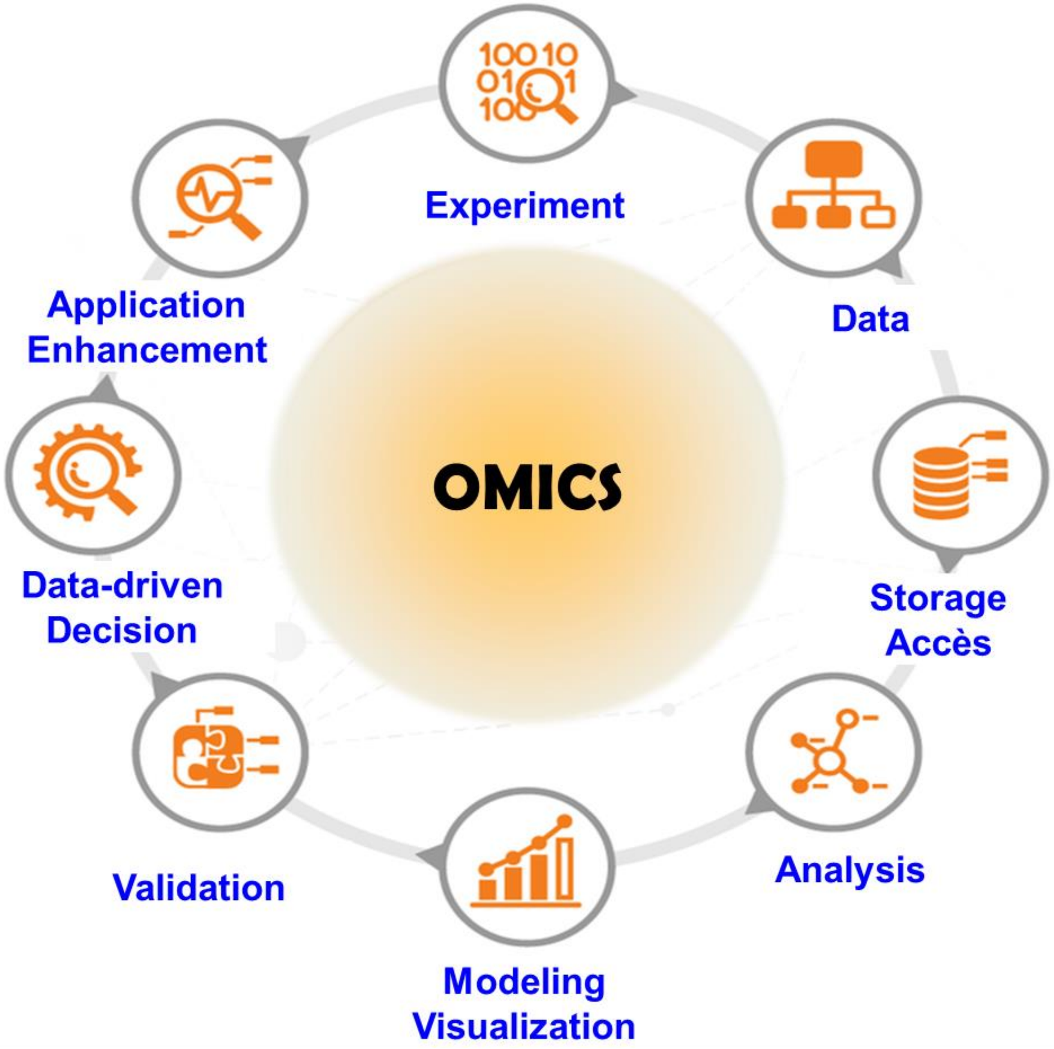


Glycan Biosynthesis and Metabolism
N-Glycan biosynthesis
O-Mannosyl glycan biosynthesis
High-mannose type N-glycan biosynthesis
Glycosphingolipid biosynthesis - globoseries
Glycosphingolipid biosynthesis - lact and neolacto series
Other glycan degradation
[B] Glycosyltransferases
O-Glycan biosynthesis
Keratan sulfate biosynthesis
Glycosaminoglycan degradation
Chondroitin sulfate biosynthesis
Lipopolysaccharide biosynthesis
Heparan sulfate biosynthesis
Biosynthesis of ansamycins
Starch and sucrose metabolism
Galactose metabolism
Polyketide sugar unit biosynthesis
Streptomycin biosynthesis
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
[B] Proteoglycans
Glycosphingolipid biosynthesis - ganglioseries
Arachidonic acid metabolism
alpha-Linolenic acid metabolism
Amino sugar and nucleotide sugar metabolism
Fructose and mannose metabolism
Pentose phosphate metabolism
Biosynthesis of 12-, 14- and 16-membered macrolides
Glycerophospholipid metabolism
Linoleic acid metabolism
Phenylpropanoid biosynthesis
[B] Lipids
Glycerolipid metabolism
Photosynthesis - antenna proteins
[B] Phytochemical compounds
Lipid Metabolism
Ether lipid metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Photosynthesis
Isoflavonoid biosynthesis
Biosynthesis of unsaturated fatty acids
Sphingolipid metabolism
Inositol phosphate metabolism
Glycolysis / Gluconeogenesis
[B] Photosynthesis Proteins
Energy Metabolism
Fatty acid biosynthesis
Fatty acid metabolism
Peptidoglycan biosynthesis
Flavonoid biosynthesis
Carotenoid biosynthesis
Biosynthesis of type II polyketide products
Aminophosphonate metabolism
Retinol metabolism
Pyruvate metabolism
D-Alanine metabolism
3-Chloroacrylic acid degradation
Flavone and flavonol biosynthesis
Anthocyanin biosynthesis
Biosynthesis of type II polyketide backbone
Synthesis and degradation of Ketone bodies
Limonene and pinene degradation
Zeatin biosynthesis
Butanoate metabolism
Fatty acid elongation in mitochondria
Selenoamino acid metabolism
Cyanoamino acid metabolism
Monoterpenoid biosynthesis
Propanoate metabolism
Lysine degradation
Sesquiterpenoid biosynthesis
Lipoic acid metabolism
Tetrachloroethene degradation
Reductive carboxylate cycle (CO2 fixation)
Terpenoid backbone biosynthesis
Caprolactam degradation
Carbon fixation
C5-Branched dibasic acid metabolism
Steroid biosynthesis
Methane metabolism
Citrate cycle (TCA cycle)
Lysine biosynthesis
Insect hormone biosynthesis
Primary bile acid biosynthesis
Atrazine degradation
Glyoxylate and dicarboxylate metabolism
Alanine, aspartate and glutamate metabolism
C21-Steroid hormone metabolism
Androgen and estrogen metabolism
Tetracycline biosynthesis
1,4-Dichlorobenzene degradation
Diterpenoid biosynthesis
Naphthalene and anthracene degradation
Biphenyl degradation
gamma-Hexachlorocyclohexane degradation
1,2-Dichloroethane degradation
Biosynthesis of Secondary Metabolism
Fluorobenzoate degradation
Styrene degradation
Bisphenol A degradation
Fluorene degradation
Oxidative phosphorylation
2,4-Dichlorobenzoate degradation
Brassinosteroid biosynthesis
Secondary bile acid biosynthesis
Toluene and xylene degradation
Ethylbenzene degradation
1,1,1-Trichloro-2,2-bis(4-chlorophenyl) ethane (DDT) degradation
Benzoate degradation via CoA ligation
Benzoate degradation via hydroxylation
Carbazole degradation
[B] Cytochrome P450 Substrates
[B] Cytochrome P450
Xenobiotics Biodegradation and Metabolism
Nucleotide Metabolism
Puromycin biosynthesis
Purine metabolism
Caffeine metabolism
Thiamine metabolism
Ascorbate and aldarate metabolism
beta-Alanine metabolism
Tyrosine metabolism
Folate biosynthesis
Pentose and glucuronate interconversions
Pyrimidine metabolism
Riboflavin metabolism
Vitamin B6 metabolism
Biosynthesis of vancomycin group antibiotics
Phenylalanine, tyrosine and tryptophan biosynthesis
Phenylalanine metabolism
Acridone alkaloid biosynthesis
One carbon pool by folate
Isoquinoline alkaloid biosynthesis
Novobiocin biosynthesis
Ubiquinone and other terpenoid-quinone biosynthesis
Metabolism of Cofactors and Vitamins
Sulfur metabolism
Taurine and hypotaurine metabolism
Porphyrin and chlorophyll metabolism
Biosynthesis of siderophore group nonribosomal peptides
Tryptophan metabolism
Indole alkaloid biosynthesis
Histidine metabolism
Benzoxazinone biosynthesis
Glycine, serine and threonine metabolism
Valine, leucine and isoleucine biosynthesis
Pantothenate and CoA biosynthesis
Urea cycle
Clavulanic acid biosynthesis
Valine, leucine and isoleucine degradation
Nitrogen metabolism
Cystein and methionine metabolism
D-Arginine and D-ornithine metabolism
Glucosinolate biosynthesis
Arginine and proline metabolism
Glutathione metabolism
Tropane, piperidine and pyridine alkaloid biosynthesis
Penicillin and cephalosporin biosynthesis
D-Glutamine and D-glutamate metabolism
Biotin metabolism
Nicotinate and nicotinamide metabolism
beta-Lactam resistance
Metabolism of Other Amino Acid

Metabolomics (Targeted)


Glycan Biosynthesis and Metabolism
N-Glycan biosynthesis
O-Mannosyl glycan biosynthesis
High-mannose type N-glycan biosynthesis
Glycosphingolipid biosynthesis - globoseries
Glycosphingolipid biosynthesis - lact and neolacto series
[B] Glycosyltransferases
Other glycan degradation
O-Glycan biosynthesis
Keratan sulfate biosynthesis
Glycosaminoglycan degradation
Chondroitin sulfate biosynthesis
Lipopolysaccharide biosynthesis
Heparan sulfate biosynthesis
Biosynthesis of ansamycins
Starch and sucrose metabolism
Galactose metabolism
Polyketide sugar unit biosynthesis
Puromycin biosynthesis
Purine metabolism
Nucleotide Metabolism
Caffeine metabolism
Thiamine metabolism
Ascorbate and aldarate metabolism
[B] Proteoglycans
Glycosphingolipid biosynthesis - ganglioseries
Arachidonic acid metabolism
alpha-Linolenic acid metabolism
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
Streptomycin biosynthesis
Pentose and glucuronate interconversions
beta-Alanine metabolism
Pyrimidine metabolism
Tyrosine metabolism
Riboflavin metabolism
Folate biosynthesis
Vitamin B6 metabolism
Biosynthesis of vancomycin group antibiotics
Amino sugar and nucleotide sugar metabolism
Pentose phosphate metabolism
Biosynthesis of 12-, 14- and 16-membered macrolides
Fructose and mannose metabolism
Glycerophospholipid metabolism
Phenylalanine, tyrosine and tryptophan biosynthesis
Phenylalanine metabolism
Linoleic acid
Photosynthesis - antenna proteins
Acridone alkaloid biosynthesis
Novobiocin biosynthesis
Isoquinoline alkaloid biosynthesis
One carbon pool by folate
[B] Phytochemical compounds
Phenylpropanoid biosynthesis
Lipid Metabolism
[B] Lipids
Glycerolipid metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Ubiquinone and other terpenoid-quinone biosynthesis
Metabolism of Cofactors and Vitamins
Isoflavonoid biosynthesis
Biosynthesis of unsaturated fatty acid
Ether lipid metabolism
Sphingolipid metabolism
Inositol phosphate metabolism
Photosynthesis
Sulfur metabolism
Taurine and hypotaurine metabolism
Porphyrin and chlorophyll metabolism
Glycolysis / Gluconeogenesis
[B] Photosynthesis Proteins
Biosynthesis of siderophore group nonribosomal peptides
Flavonoid biosynthesis
Carotenoid biosynthesis
Fatty acid biosynthesis
Fatty acid metabolism
Peptidoglycan biosynthesis
Energy Metabolism
Tryptophan metabolism
Indole alkaloid biosynthesis
Retinol metabolism
Biosynthesis of type II polyketide products
Aminophosphonate metabolism
Histidine metabolism
D-Alanine metabolism
3-Chloroacrylic acid degradation
Pyruvate metabolism
Benzoxazinone biosynthesis
Glycine, serine and threonine metabolism
Valine, leucine and isoleucine biosynthesis
Pantothenate and CoA biosynthesis
Flavone and flavonol biosynthesis
Anthocyanin biosynthesis
Biosynthesis of type II polyketide backbone
Synthesis and degradation of Ketone bodies
Limonene and pinene degradation
Zeatin biosynthesis
Selenoamino acid metabolism
Cyanoamino acid metabolism
Urea cycle
Clavulanic acid biosynthesis
Monoterpenoid biosynthesis
Butanoate metabolism
Fatty acid elongation in mitochondria
Propanoate metabolism
Lysine degradation
Valine, leucine and isoleucine degradation
Sesquiterpenoid biosynthesis
Lipoic acid metabolism
Tetrachloroethene degradation
Reductive carboxylate cycle (CO2 fixation)
Terpenoid backbone biosynthesis
Caprolactam degradation
Carbon fixation
C5-Branched dibasic acid metabolism
Lysine biosynthesis
Cystein and methionine metabolism
Nitrogen metabolism
D-Arginine and D-ornithine metabolism
Steroid biosynthesis
Methane metabolism
Citrate cycle (TCA cycle)
Primary bile acid biosynthesis
Atrazine degradation
Alanine, aspartate and glutamate metabolism
Insect hormone biosynthesis
Glyoxylate and dicarboxylate metabolism
C21-Steroid hormone metabolism
Androgen and estrogen metabolism
Tetracycline biosynthesis
Glucosinolate biosynthesis
1,4-Dichlorobenzene degradation
Diterpenoid biosynthesis
Naphthalene and anthracene degradation
Biphenyl degradation
Arginine and proline metabolism
gamma-Hexachlorocyclohexane degradation
Biosynthesis of Secondary Metabolism
Fluorobenzoate degradation
1,2-Dichloroethane degradation
Bisphenol A degradation
Glutathione metabolism
Tropane, piperidine and pyridine alkaloid biosynthesis
Fluorene degradation
Styrene degradation
Oxidative phosphorylation
Penicillin and cephalosporin biosynthesis
D-Glutamine and D-glutamate metabolism
Biotin metabolism
Nicotinate and nicotinamide metabolism
Brassinosteroid biosynthesis
Secondary bile acid biosynthesis
2,4-Dichlorobenzoate degradation
Toluene and xylene degradation
Ethylbenzene degradation
1,1,1-Trichloro-2,2-bis(4-chlorophenyl) ethane (DDT) degradation
beta-Lactam resistance
Metabolism of Other Amino Acid
Benzoate degradation via CoA ligation
Benzoate degradation via hydroxylation
Carbazole degradation
Xenobiotics Biodegradation and Metabolism
[B] Cytochrome P450 Substrates
[B] Cytochrome P450

Biological Information Generation



Experimental Design

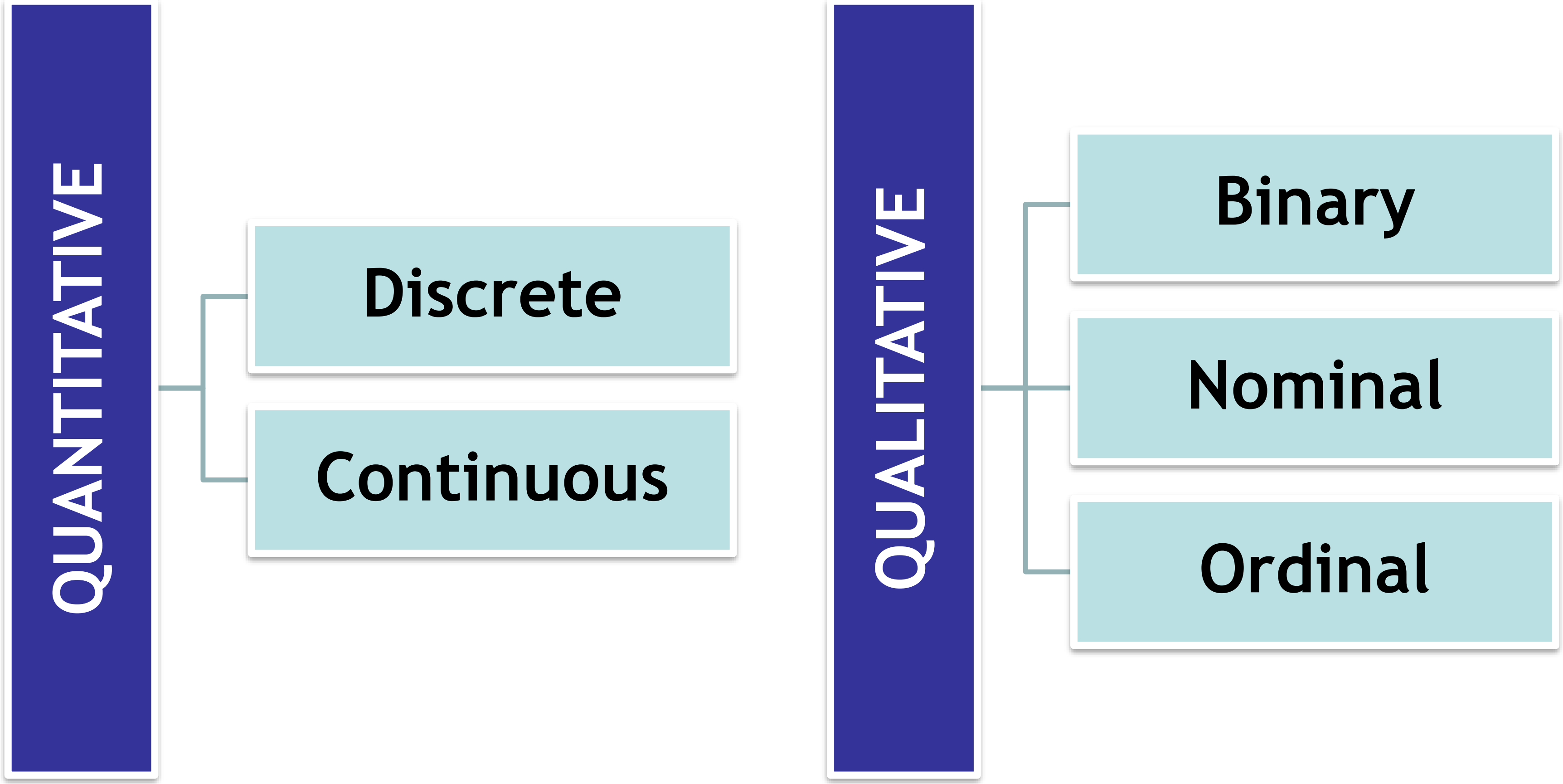
"To consult the statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of."

Sir Ronald Fisher (1938)

Experimental Design may help

1. Clear and precise study objective
2. Sample type and size
3. Sampling and sample preparation strategy
4. Number of samples / Biological - Analytical replicates
5. Analytical technology(ies)
6. Collection of meta-data (Categorical, continuous, ordinal…)
7. Confounding factors
8. Randomization
9. Data analysis strategies (univariate *vs*. multivariate)
10. Biological Interpretation and insights
11. Validation (Biological/Analytical)
12. …Name it

Know your data



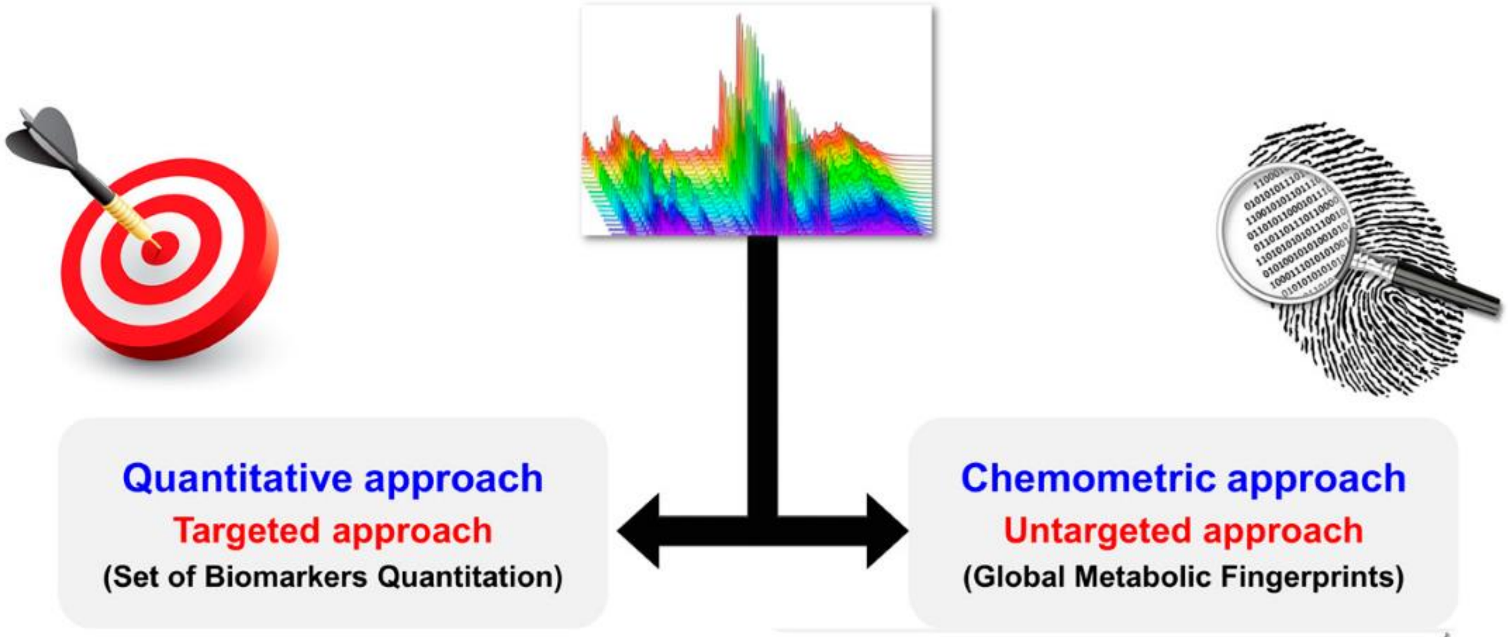
QUANTITATIVE

- Discrete
- Continuous

QUALITATIVE

- Binary
- Nominal
- Ordinal

Metabolomics



Quantitative approach

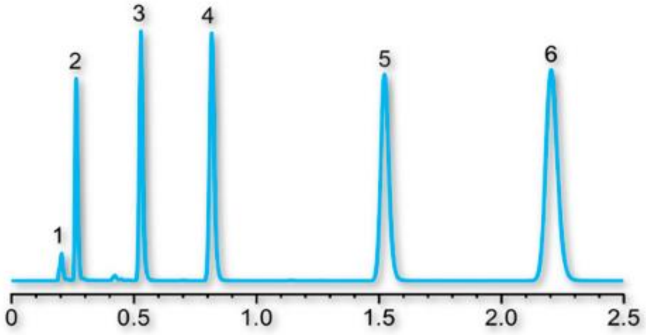
Targeted approach

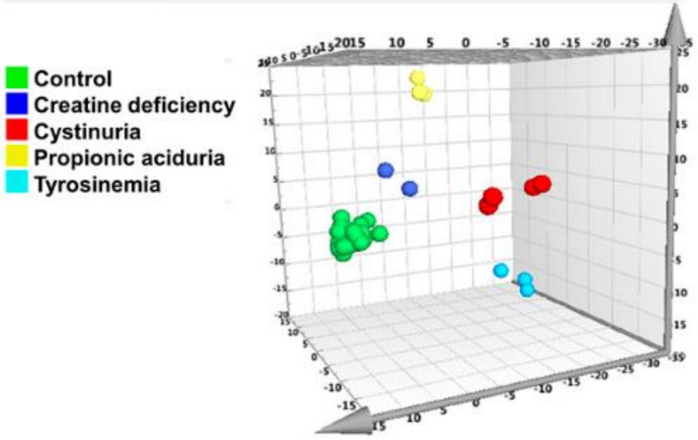
(Set of Biomarkers Quantitation)

Chemometric approach

Untargeted approach

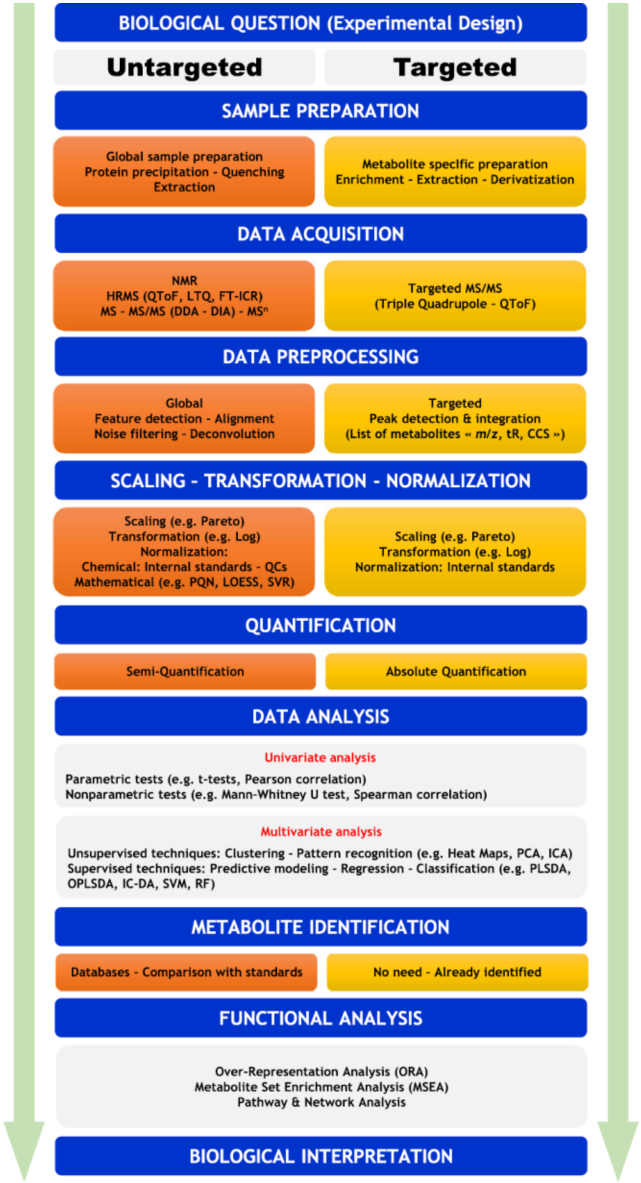
(Global Metabolic Fingerprints)





Tebani et al. IJMS. 2016

Metabolomics Workflows Overview



BIOLOGICAL QUESTION (Experimental Design)

Untargeted	Targeted
SAMPLE PREPARATION	
Global sample preparation Protein precipitation - Quenching Extraction	Metabolite specific preparation Enrichment - Extraction - Derivatization
DATA ACQUISITION	
NMR HRMS (QToF, LTQ, FT-ICR) MS - MS/MS (DDA - DIA) - MS^n	Targeted MS/MS (Triple Quadrupole - QToF)
DATA PREPROCESSING	
Global Feature detection - Alignment Noise filtering - Deconvolution	Targeted Peak detection & integration (List of metabolites « *m/z*, tR, CCS »)
SCALING - TRANSFORMATION - NORMALIZATION	
Scaling (e.g. Pareto) Transformation (e.g. Log) Normalization: Chemical: Internal standards - QCs Mathematical (e.g. PQN, LOESS, SVR)	Scaling (e.g. Pareto) Transformation (e.g. Log) Normalization: Internal standards
QUANTIFICATION	
Semi-Quantification	Absolute Quantification

DATA ANALYSIS

Univariate analysis

Parametric tests (e.g. t-tests, Pearson correlation)
Nonparametric tests (e.g. Mann-Whitney U test, Spearman correlation)

Multivariate analysis

Unsupervised techniques: Clustering - Pattern recognition (e.g. Heat Maps, PCA, ICA)
Supervised techniques: Predictive modeling - Regression - Classification (e.g. PLSDA, OPLSDA, IC-DA, SVM, RF)

METABOLITE IDENTIFICATION

Untargeted	Targeted
Databases - Comparison with standards	No need - Already identified

FUNCTIONAL ANALYSIS

Over-Representation Analysis (ORA)
Metabolite Set Enrichment Analysis (MSEA)
Pathway & Network Analysis

BIOLOGICAL INTERPRETATION

Tebani et al. JIMD. 2017

Metabolomics worflow

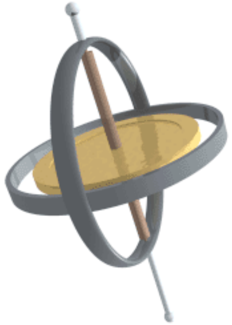
1

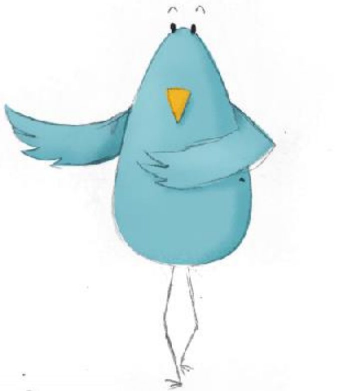
2

3

Tebani, A. *IJMS.* 2016, *17*, 1167

Biological Information Extraction





NMR spectroscopy

Mass spectrometry

Biological Information Extraction

NMR spectroscopy

Tissues, biofluids and extracts

Interaction of spin active nuclei (^{1}H, ^{13}C, ^{31}P) with electromagnetic fields gives molecular information

Non-destructive

Cross-instrument robustness

Mass spectrometry

Tissues, biofluids and extracts

Mass to charge ratio *(m/z)*

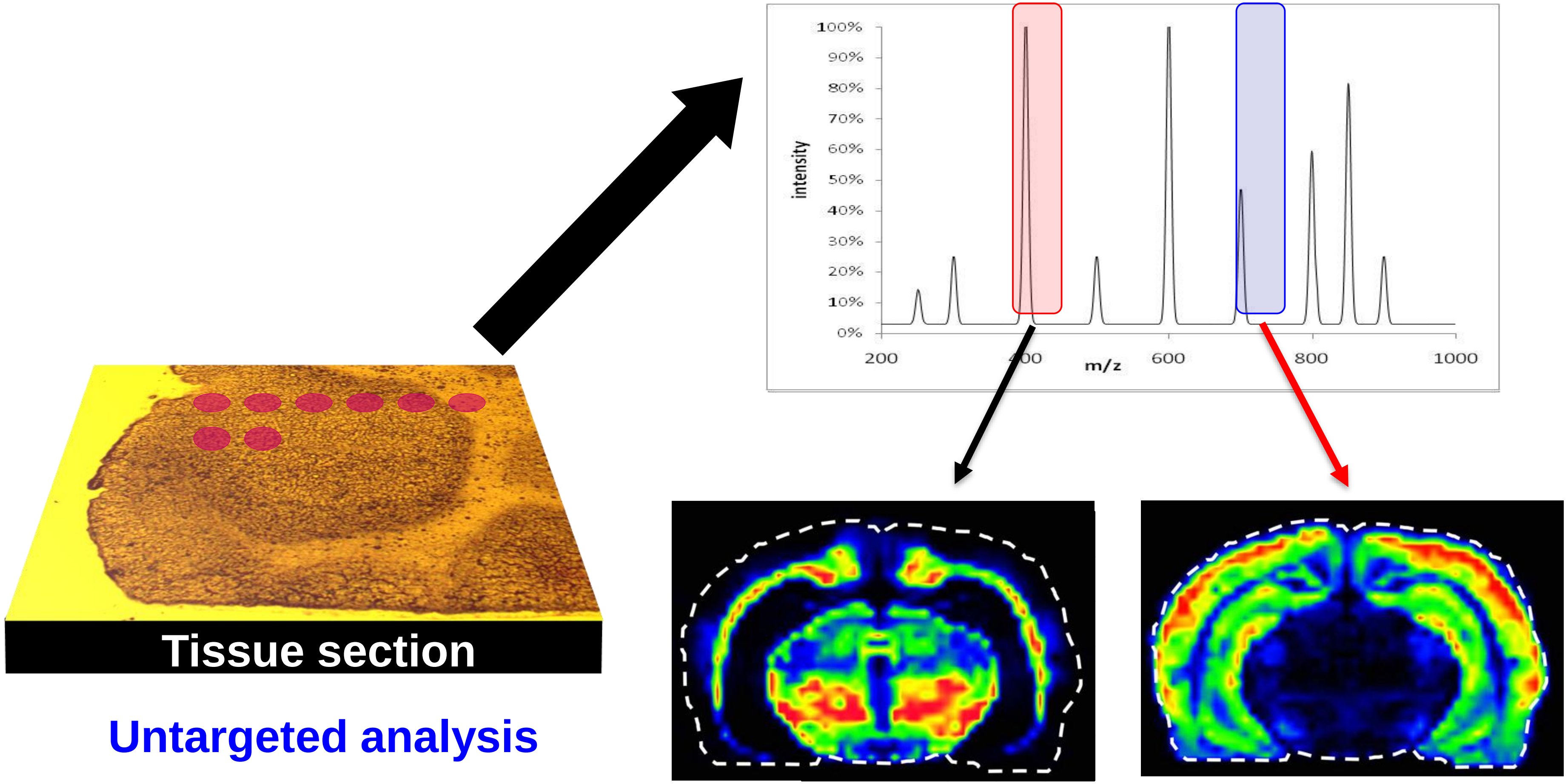
Sensitivity

Higher metabolome coverage

Tebani et al. JIMD. 2016

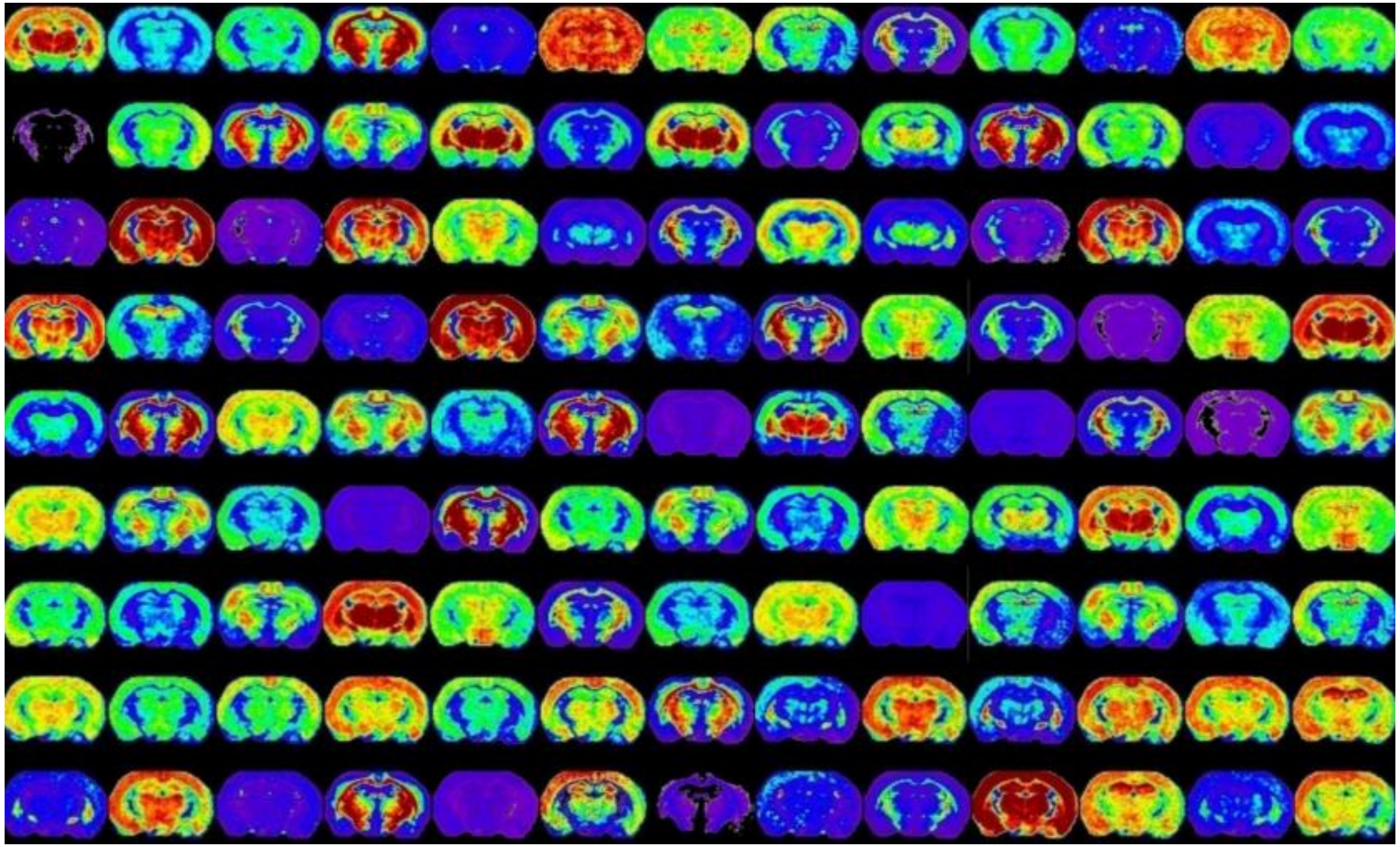
Metabolomics-based imaging

Metabolomic Imaging



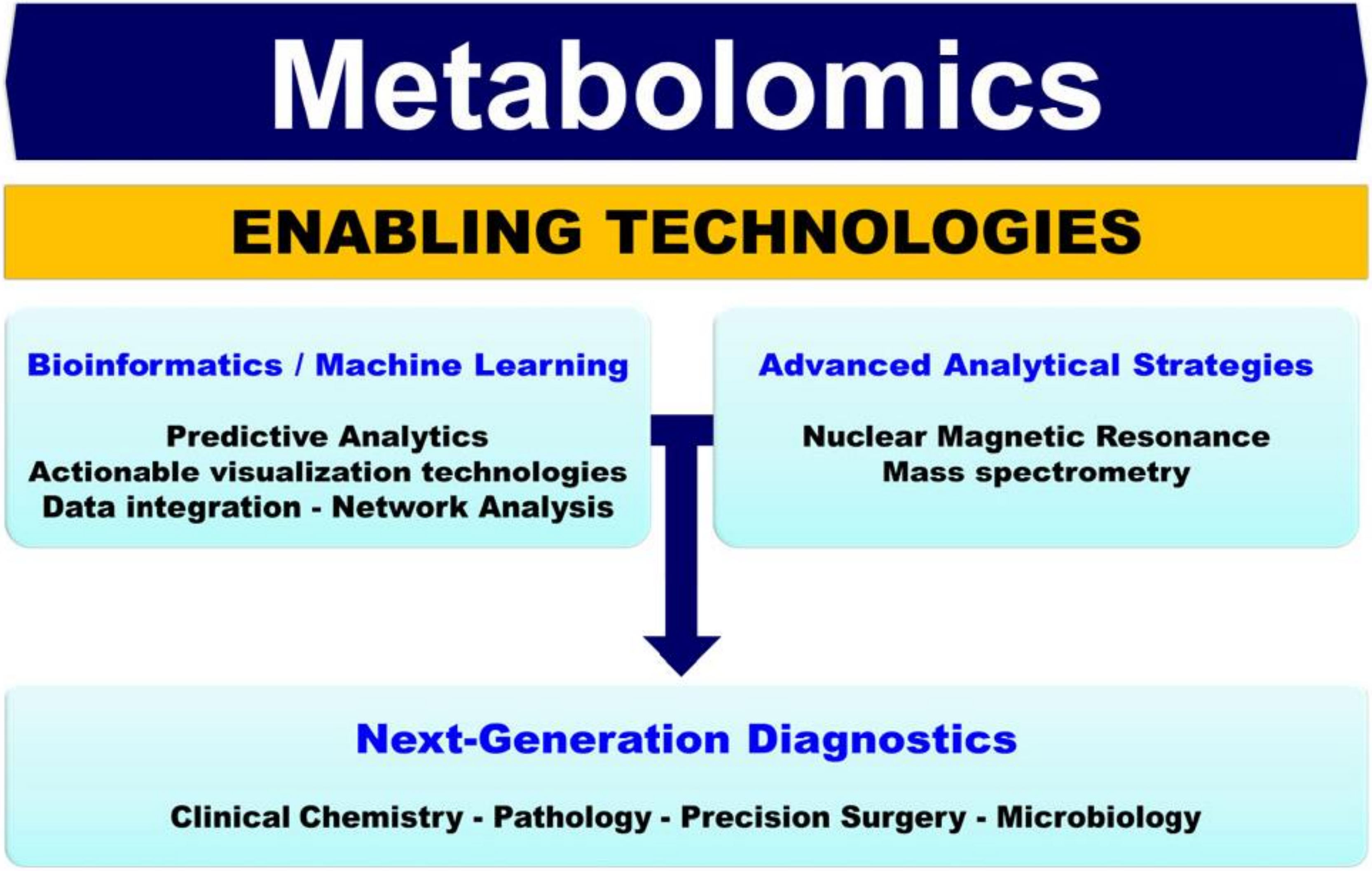
Untargeted analysis

Metabolomics-based imaging



Journal of Mass Spectrometry
Volume 46, Issue 2, pages 209-222, 24 JAN 2011 DOI: 10.1002/jms.1876
http://onlinelibrary.wiley.com/doi/10.1002/jms.1876/full#fig1

Applications



Metabolomics

ENABLING TECHNOLOGIES

Bioinformatics / Machine Learning

Predictive Analytics
Actionable visualization technologies
Data integration - Network Analysis

Advanced Analytical Strategies

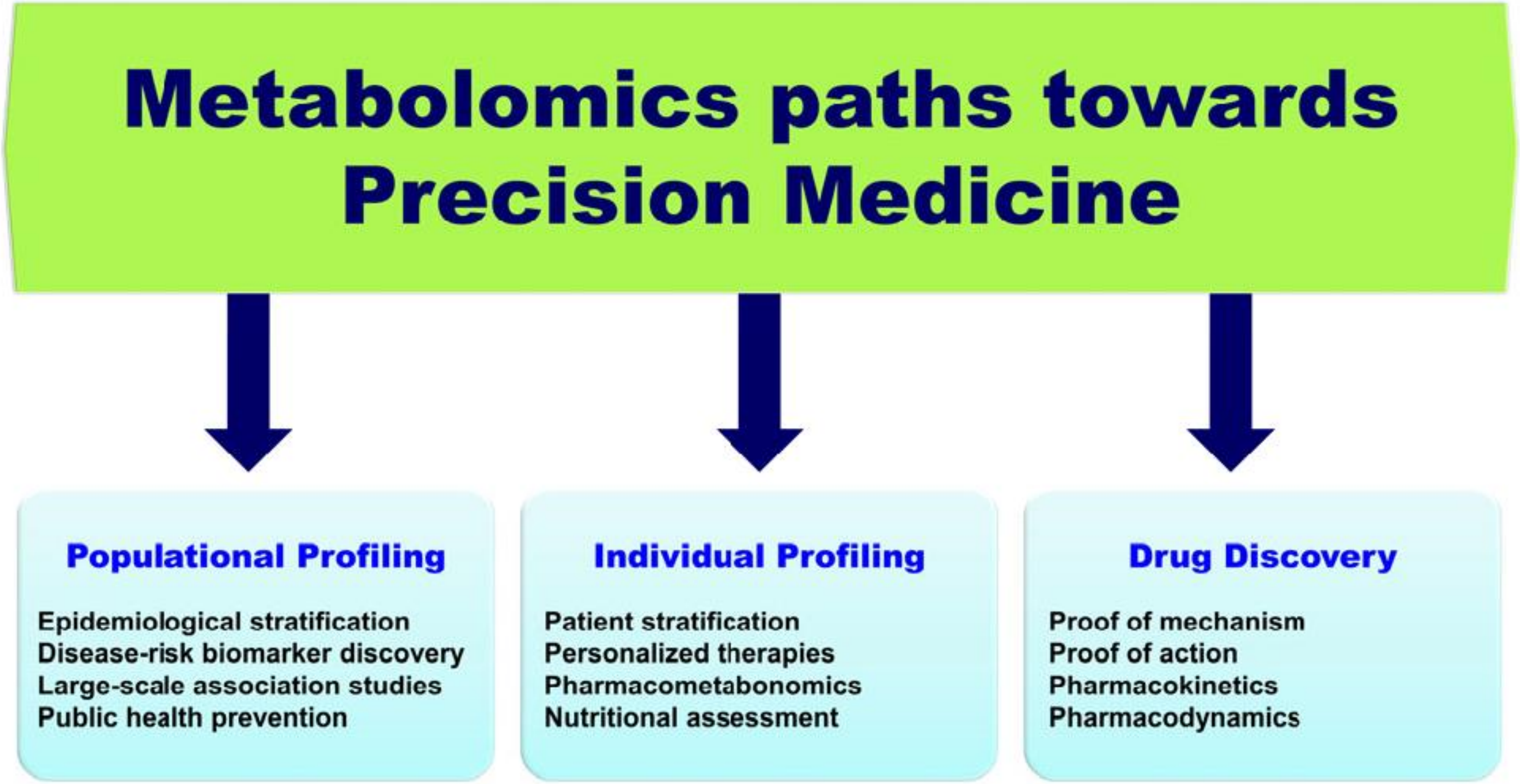
Nuclear Magnetic Resonance
Mass spectrometry

Next-Generation Diagnostics

Clinical Chemistry - Pathology - Precision Surgery - Microbiology

Bekri S, *Expert Review of Precision Medicine and Drug Development* 1.6 (2016): 517-532
https://doi.org/10.1080/23808993.2016.1273067

Applications



Metabolomics paths towards Precision Medicine

Populational Profiling

Epidemiological stratification
Disease-risk biomarker discovery
Large-scale association studies
Public health prevention

Individual Profiling

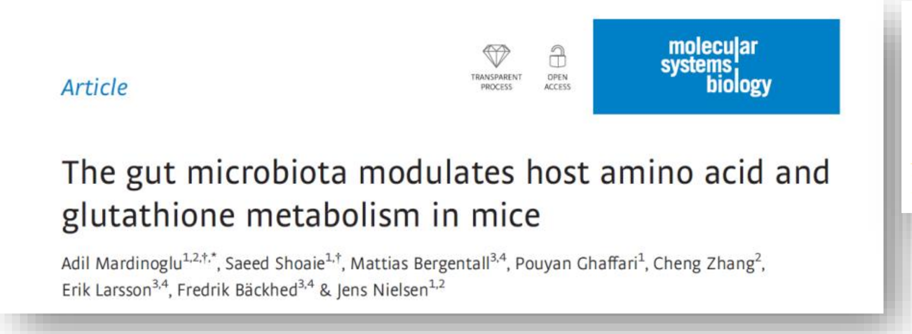
Patient stratification
Personalized therapies
Pharmacometabonomics
Nutritional assessment

Drug Discovery

Proof of mechanism
Proof of action
Pharmacokinetics
Pharmacodynamics

Bekri S, *Expert Review of Precision Medicine and Drug Development* 1.6 (2016): 517-532
https://doi.org/10.1080/23808993.2016.1273067

Applications



TRANSPARENT PROCESS
OPEN ACCESS
molecular systems biology

Article

The gut microbiota modulates host amino acid and glutathione metabolism in mice

Adil Mardinoglu[1,2,†,*], Saeed Shoaie[1,†], Mattias Bergentall[3,4], Pouyan Ghaffari[1], Cheng Zhang[2], Erik Larsson[3,4], Fredrik Bäckhed[3,4] & Jens Nielsen[1,2]

Host-Microbiota intecations

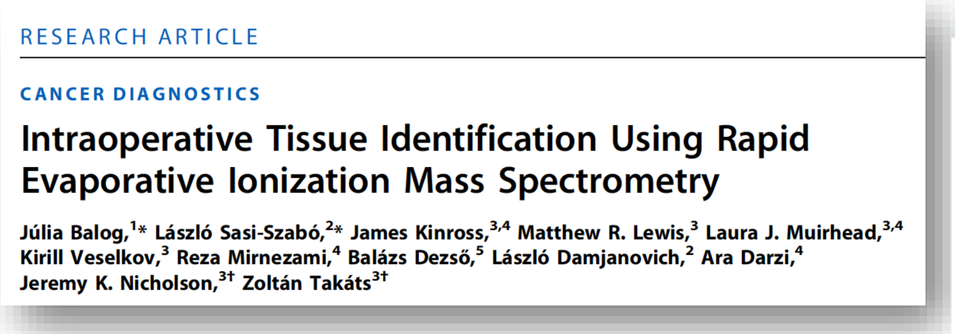
Int. J. Mol. Sci. **2016**, *17*(7), 1167; doi:10.3390/ijms17071167 Open Access

Review

Clinical Metabolomics: The New Metabolic Window for Inborn Errors of Metabolism Investigations in the Post-Genomic Era

Abdellah Tebani [1,2,3] ✉, Lenaig Abily-Donval [2,4] ✉, Carlos Afonso [3] ✉, Stéphane Marret [2,4] ✉ and Soumeya Bekri [1,2,*] ✉

Inherited Metabolic Diseases



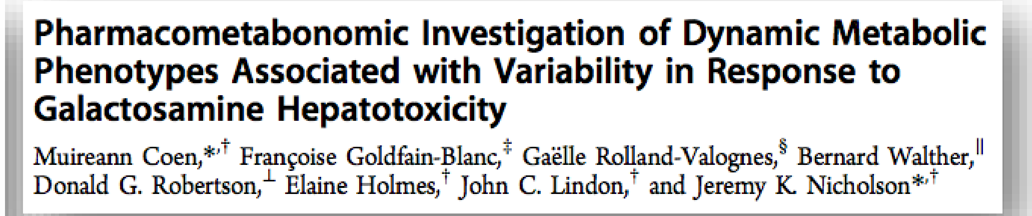
RESEARCH ARTICLE

CANCER DIAGNOSTICS

Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry

Júlia Balog,[1]* László Sasi-Szabó,[2]* James Kinross,[3,4] Matthew R. Lewis,[3] Laura J. Muirhead,[3,4] Kirill Veselkov,[3] Reza Mirnezami,[4] Balázs Dezső,[5] László Damjanovich,[2] Ara Darzi,[4] Jeremy K. Nicholson,[3†] Zoltán Takáts[3†]

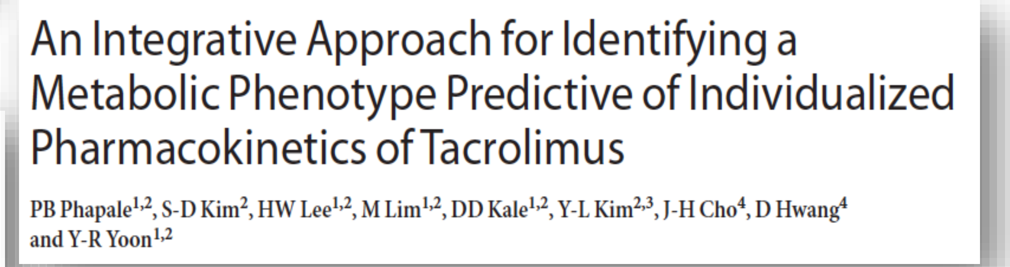
Pathology and Cancer



Pharmacometabonomic Investigation of Dynamic Metabolic Phenotypes Associated with Variability in Response to Galactosamine Hepatotoxicity

Muireann Coen,[*,†] Françoise Goldfain-Blanc,[‡] Gaëlle Rolland-Valognes,[§] Bernard Walther,[‖] Donald G. Robertson,[⊥] Elaine Holmes,[†] John C. Lindon,[†] and Jeremy K. Nicholson[*,†]

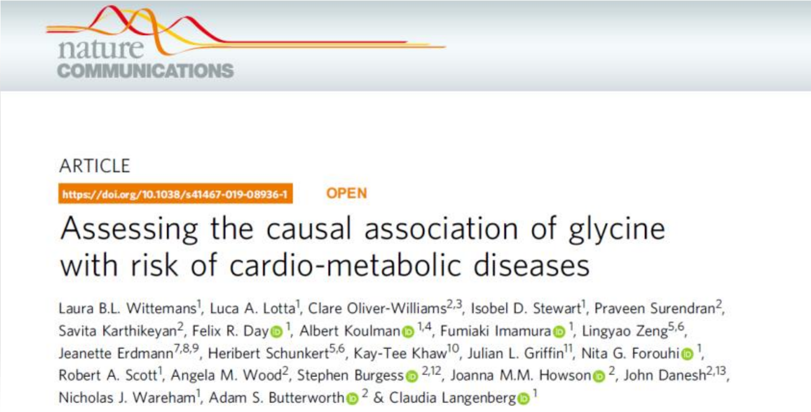
Responder Non-responder Prediction



An Integrative Approach for Identifying a Metabolic Phenotype Predictive of Individualized Pharmacokinetics of Tacrolimus

PB Phapale[1,2], S-D Kim[2], HW Lee[1,2], M Lim[1,2], DD Kale[1,2], Y-L Kim[2,3], J-H Cho[4], D Hwang[4] and Y-R Yoon[1,2]

Human Drug Pharmacokinetics



nature COMMUNICATIONS

ARTICLE

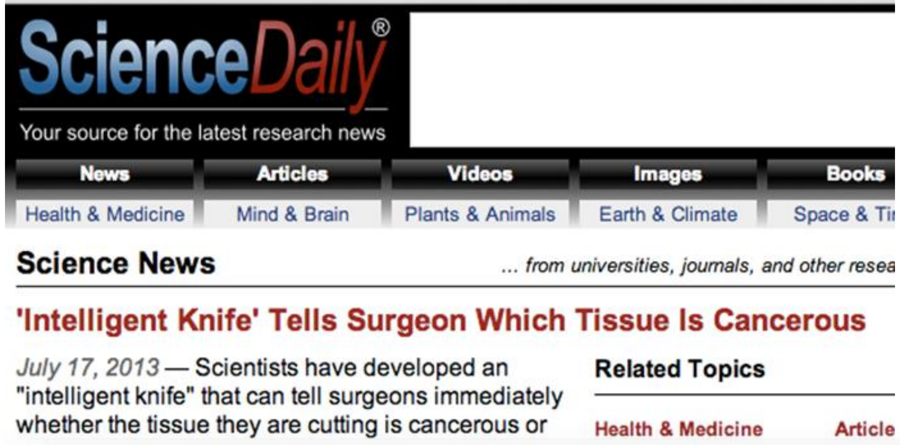
https://doi.org/10.1038/s41467-019-08936-1 OPEN

Assessing the causal association of glycine with risk of cardio-metabolic diseases

Laura B.L. Wittemans[1], Luca A. Lotta[1], Clare Oliver-Williams[2,3], Isobel D. Stewart[1], Praveen Surendran[2], Savita Karthikeyan[2], Felix R. Day[1], Albert Koulman[1,4], Fumiaki Imamura[1], Lingyao Zeng[5,6], Jeanette Erdmann[7,8,9], Heribert Schunkert[5,6], Kay-Tee Khaw[10], Julian L. Griffin[11], Nita G. Forouhi[1], Robert A. Scott[1], Angela M. Wood[2], Stephen Burgess[2,12], Joanna M.M. Howson[2], John Danesh[2,13], Nicholas J. Wareham[1], Adam S. Butterworth[2] & Claudia Langenberg[1]

Large-scale epidemiological profiling

Metabolomics-based imaging



ScienceDaily®

Your source for the latest research news

News | Articles | Videos | Images | Books

Health & Medicine | Mind & Brain | Plants & Animals | Earth & Climate | Space & Ti

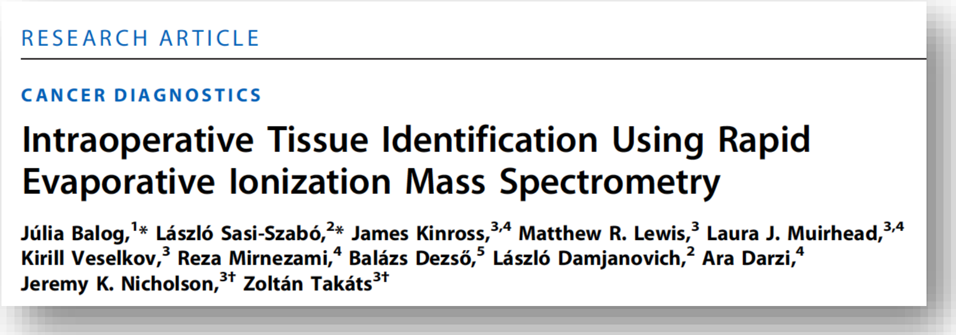
Science News ... from universities, journals, and other resea

'Intelligent Knife' Tells Surgeon Which Tissue Is Cancerous

July 17, 2013 — Scientists have developed an "intelligent knife" that can tell surgeons immediately whether the tissue they are cutting is cancerous or

Related Topics

Health & Medicine | Article

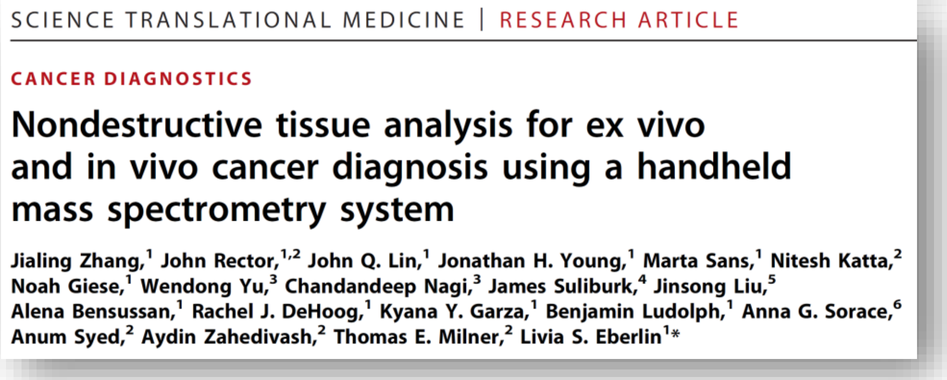


RESEARCH ARTICLE

CANCER DIAGNOSTICS

Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry

Júlia Balog,[1*] László Sasi-Szabó,[2*] James Kinross,[3,4] Matthew R. Lewis,[3] Laura J. Muirhead,[3,4] Kirill Veselkov,[3] Reza Mirnezami,[4] Balázs Dezső,[5] László Damjanovich,[2] Ara Darzi,[4] Jeremy K. Nicholson,[3†] Zoltán Takáts[3†]



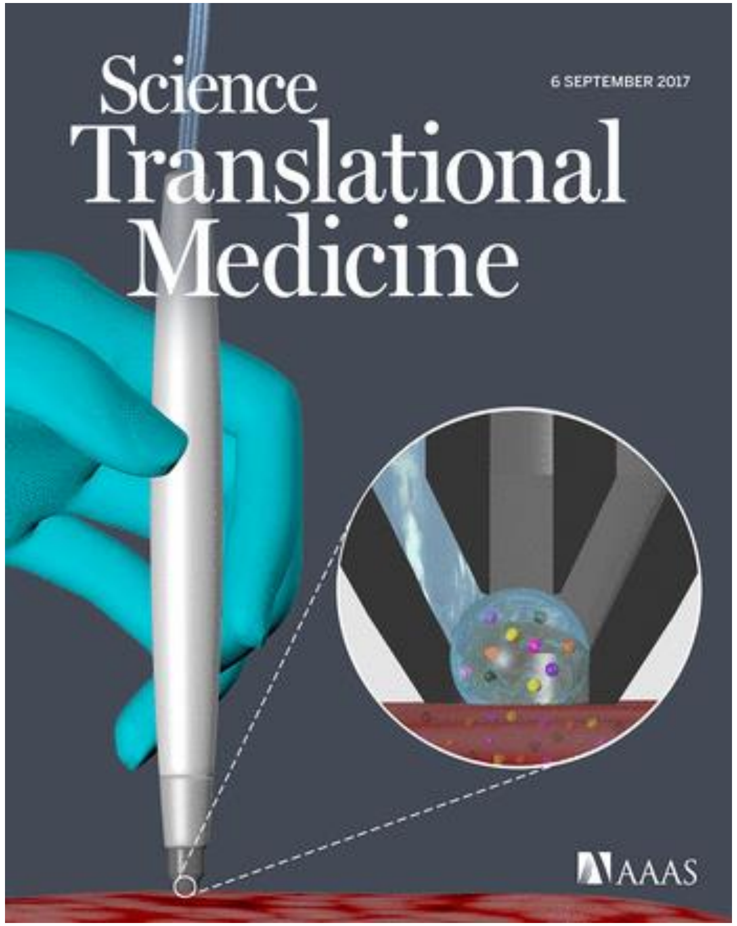
SCIENCE TRANSLATIONAL MEDICINE | RESEARCH ARTICLE

CANCER DIAGNOSTICS

Nondestructive tissue analysis for ex vivo and in vivo cancer diagnosis using a handheld mass spectrometry system

Jialing Zhang,[1] John Rector,[1,2] John Q. Lin,[1] Jonathan H. Young,[1] Marta Sans,[1] Nitesh Katta,[2] Noah Giese,[1] Wendong Yu,[3] Chandandeep Nagi,[3] James Suliburk,[4] Jinsong Liu,[5] Alena Bensussan,[1] Rachel J. DeHoog,[1] Kyana Y. Garza,[1] Benjamin Ludolph,[1] Anna G. Sorace,[6] Anum Syed,[2] Aydin Zahedivash,[2] Thomas E. Milner,[2] Livia S. Eberlin[1*]

Real time Metabolomics: Precision Surgery

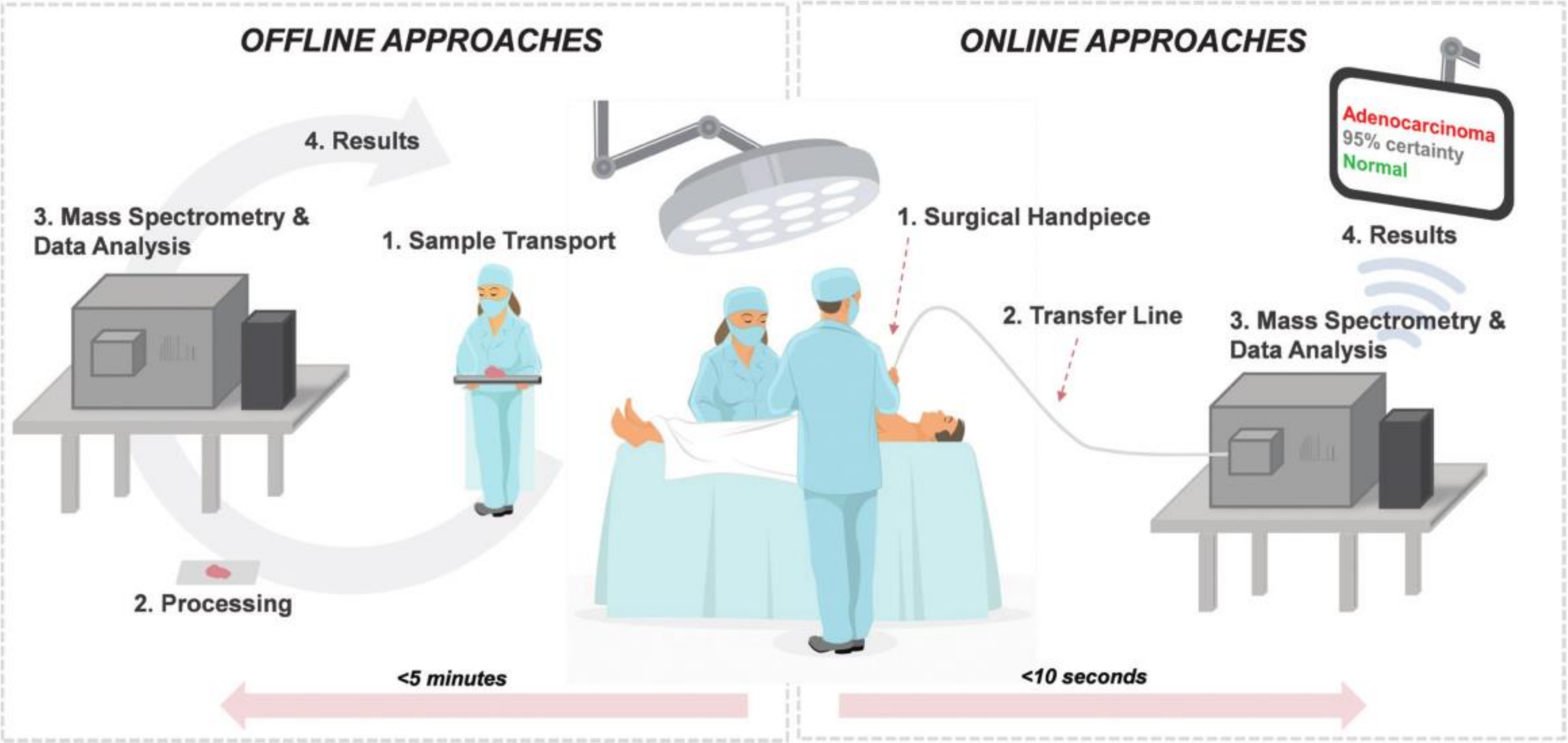


Metabolomics-based imaging

Sci Transl Med 17 July 2013: Vol. 5, Issue 194, p. 194ra93

Metabolomics-based imaging

Real time Metabolomics: Precision Surgery



Demian R. Ifa *et al*. Clin Chem 62:1 (2016)

Sci Transl Med 17 July 2013: Vol. 5, Issue 194, p. 194ra93

Limits

Metabolite Identification are the main bottlenecks of metabolomics for large adoption in both translational and clinical context.

Lack of standardized **annotation** of the metabolome is important for functional analysis and integration with other omics through GEMs

More **absolute quantification** of metabolites is needed (targeted and untargeted) to achieve reliability and robustness

Standardization and Harmonization is a prerequisite for large adoption

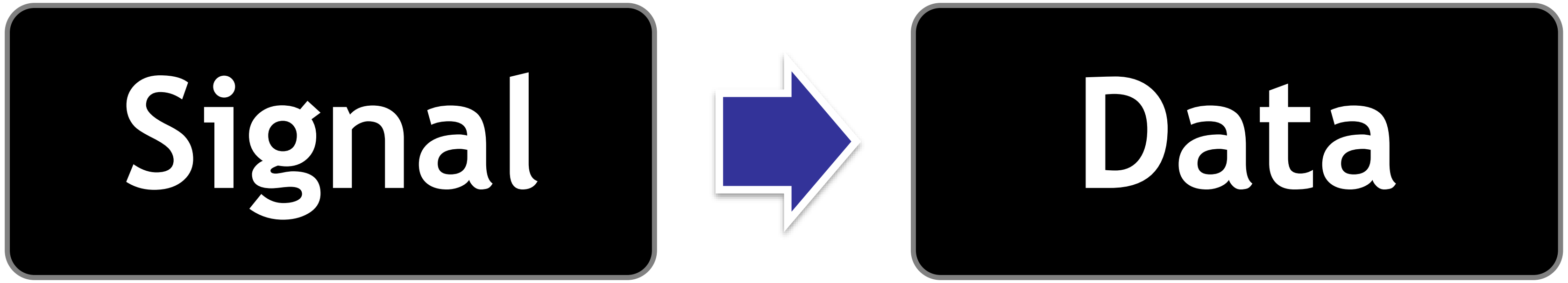
Miniaturization will enhance high-throuput

Automation, Data Visualization and Clinical Actionability at different stages, instrument-, pre- and post-analytic levels including data processing, integration and interpretation are very important issues for large clinical adoption of any diagnostic innovation

Data Analysis
From signals to numbers

PREPROCESSING

Data processing



Signal

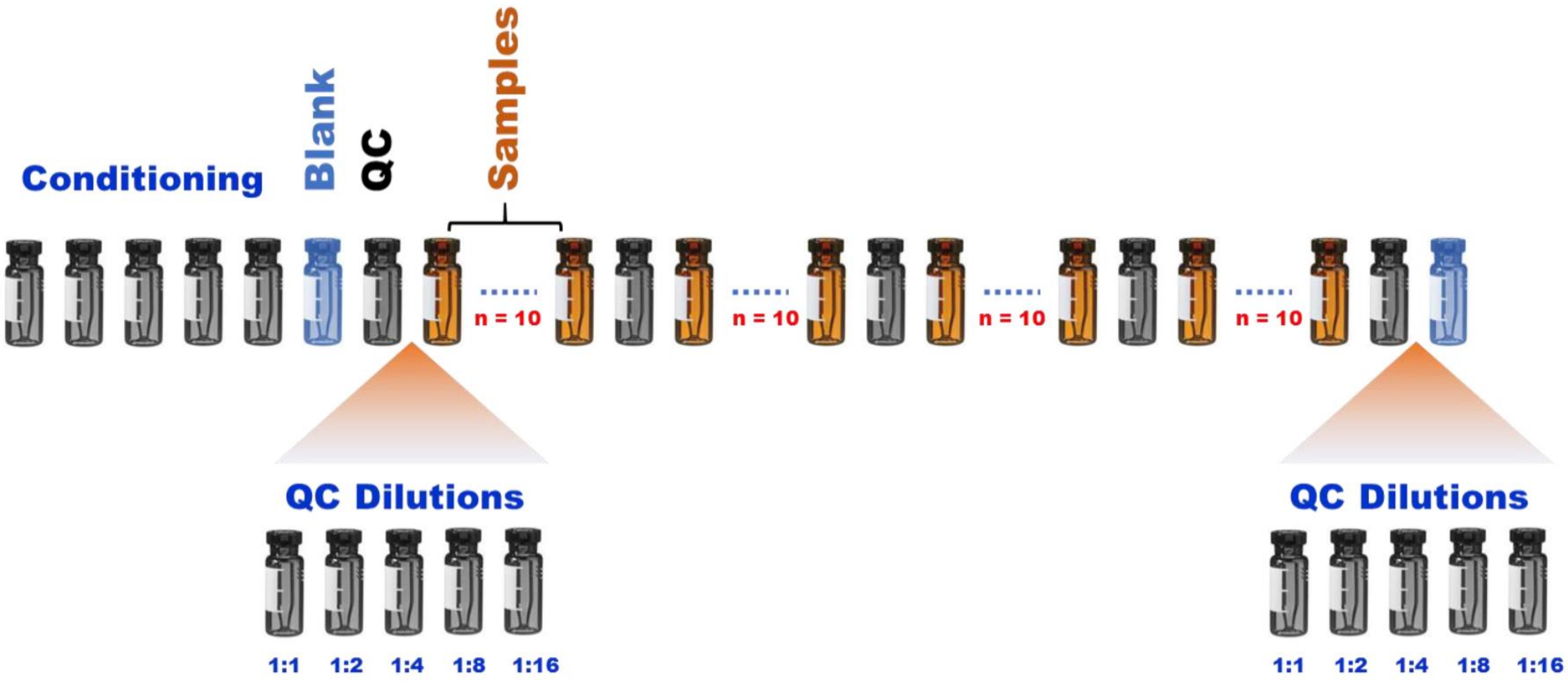
Data

Variables

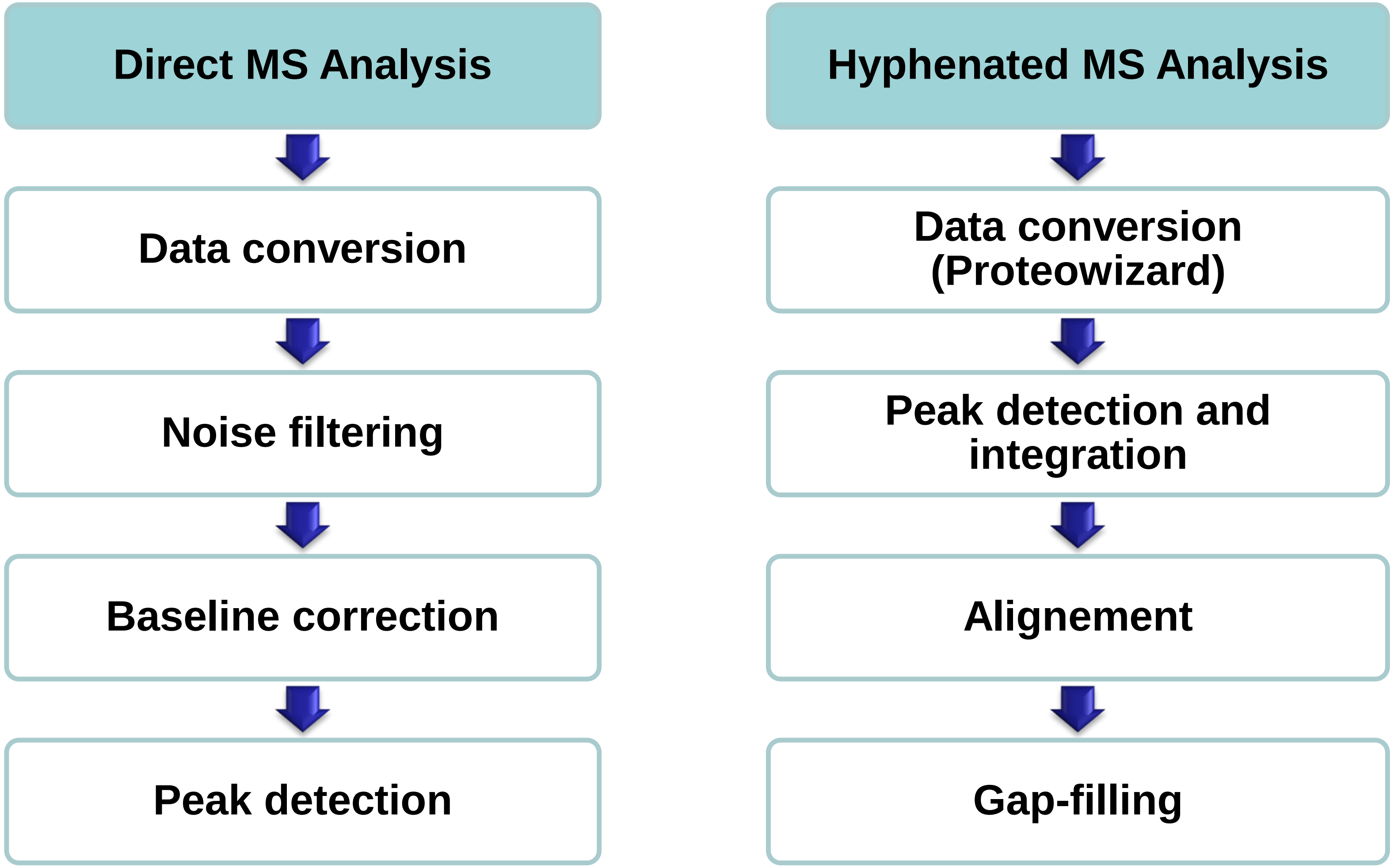
Samples

6	0	4	8	0	2	2	7	2	2
0	1	8	0	8	0	0	3	4	9
4	4	0	4	0	1	4	0	1	8
8	0	2	8	2	4	8	4		0
0	4	6	0	5	2	0	8	0	2
2	8	9	5	5	5	2	0	4	6
4	9	5	6	5	4	5	2	8	9
1	8	0	8	5	0	1	8	0	8
4	0	4	0	0	4	2	0	4	0
0	2	8	2	5	8	0	2	8	2

Data processing



Data processing



Direct MS Analysis

Data conversion

Noise filtering

Baseline correction

Peak detection

Hyphenated MS Analysis

Data conversion (Proteowizard)

Peak detection and integration

Alignement

Gap-filling

Data processing softwares

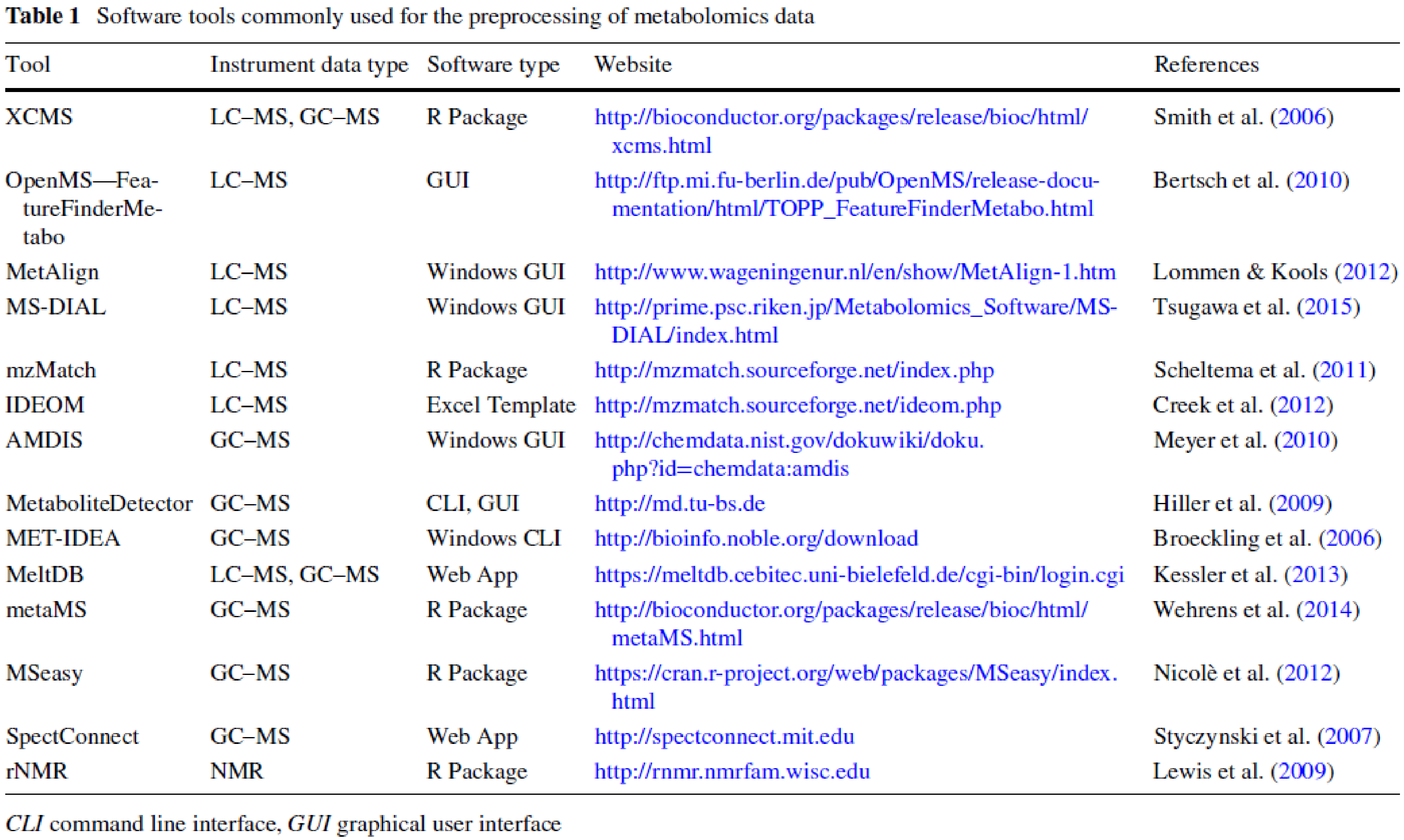


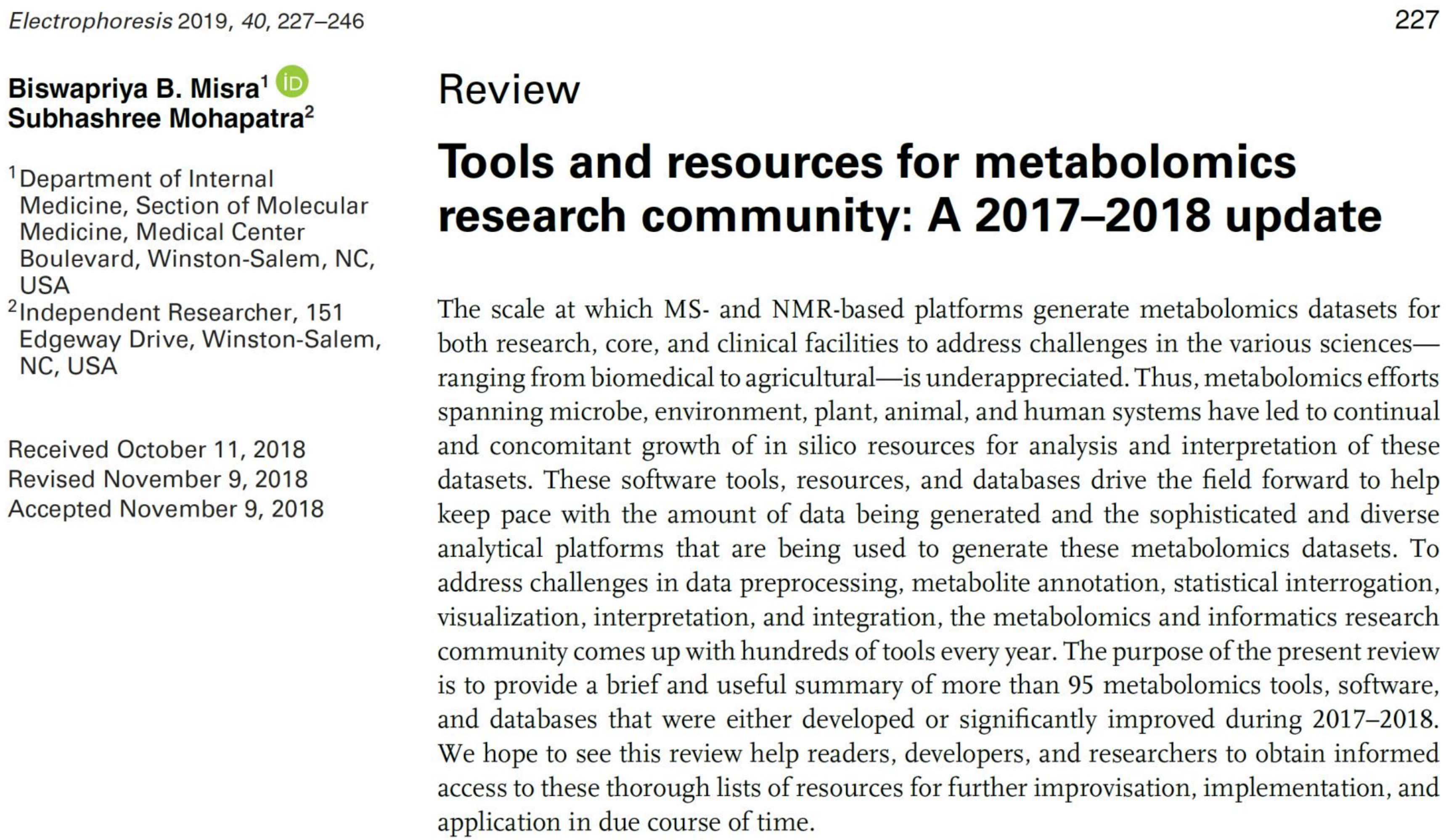
Table 1 Software tools commonly used for the preprocessing of metabolomics data

Tool	Instrument data type	Software type	Website	References
XCMS	LC–MS, GC–MS	R Package	http://bioconductor.org/packages/release/bioc/html/xcms.html	Smith et al. (2006)
OpenMS—FeatureFinderMetabo	LC–MS	GUI	http://ftp.mi.fu-berlin.de/pub/OpenMS/release-documentation/html/TOPP_FeatureFinderMetabo.html	Bertsch et al. (2010)
MetAlign	LC–MS	Windows GUI	http://www.wageningenur.nl/en/show/MetAlign-1.htm	Lommen & Kools (2012)
MS-DIAL	LC–MS	Windows GUI	http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/index.html	Tsugawa et al. (2015)
mzMatch	LC–MS	R Package	http://mzmatch.sourceforge.net/index.php	Scheltema et al. (2011)
IDEOM	LC–MS	Excel Template	http://mzmatch.sourceforge.net/ideom.php	Creek et al. (2012)
AMDIS	GC–MS	Windows GUI	http://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:amdis	Meyer et al. (2010)
MetaboliteDetector	GC–MS	CLI, GUI	http://md.tu-bs.de	Hiller et al. (2009)
MET-IDEA	GC–MS	Windows CLI	http://bioinfo.noble.org/download	Broeckling et al. (2006)
MeltDB	LC–MS, GC–MS	Web App	https://meltdb.cebitec.uni-bielefeld.de/cgi-bin/login.cgi	Kessler et al. (2013)
metaMS	GC–MS	R Package	http://bioconductor.org/packages/release/bioc/html/metaMS.html	Wehrens et al. (2014)
MSeasy	GC–MS	R Package	https://cran.r-project.org/web/packages/MSeasy/index.html	Nicolè et al. (2012)
SpectConnect	GC–MS	Web App	http://spectconnect.mit.edu	Styczynski et al. (2007)
rNMR	NMR	R Package	http://rnmr.nmrfam.wisc.edu	Lewis et al. (2009)

CLI command line interface, *GUI* graphical user interface

Spicer, R., Salek, R.M., Moreno, P. et al. Metabolomics (2017) 13: 106

Data processing softwares



Electrophoresis 2019, *40*, 227–246

227

Biswapriya B. Misra[1]
Subhashree Mohapatra[2]

[1] Department of Internal Medicine, Section of Molecular Medicine, Medical Center Boulevard, Winston-Salem, NC, USA
[2] Independent Researcher, 151 Edgeway Drive, Winston-Salem, NC, USA

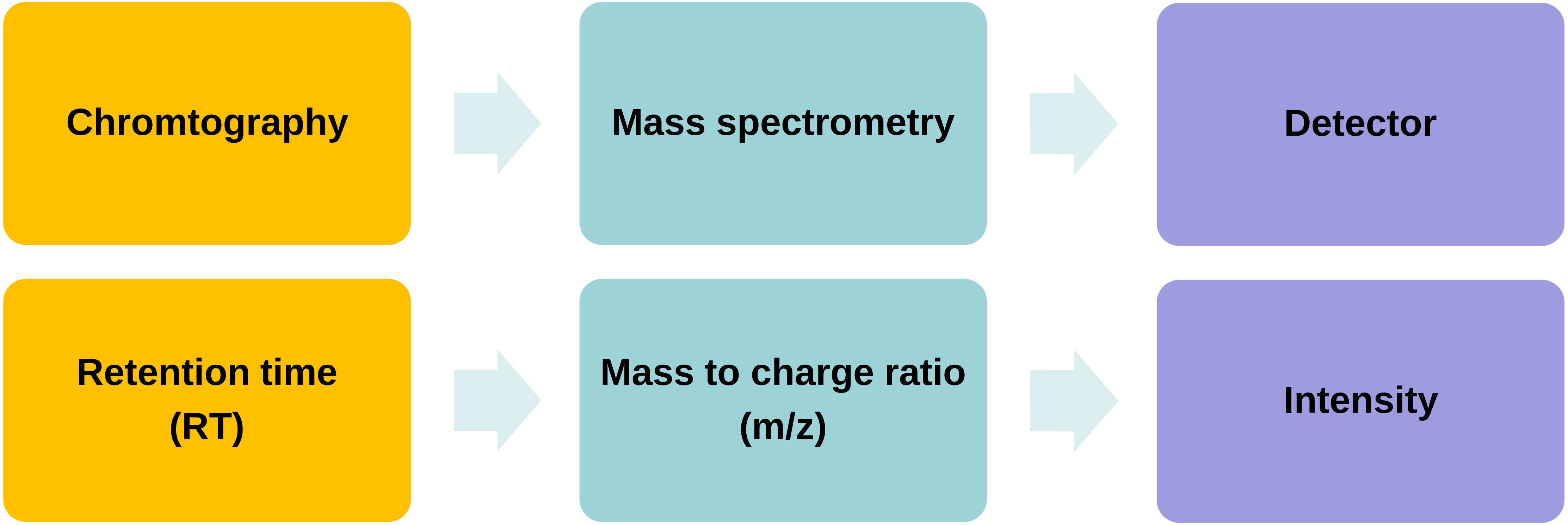
Received October 11, 2018
Revised November 9, 2018
Accepted November 9, 2018

Review

Tools and resources for metabolomics research community: A 2017–2018 update

The scale at which MS- and NMR-based platforms generate metabolomics datasets for both research, core, and clinical facilities to address challenges in the various sciences—ranging from biomedical to agricultural—is underappreciated. Thus, metabolomics efforts spanning microbe, environment, plant, animal, and human systems have led to continual and concomitant growth of in silico resources for analysis and interpretation of these datasets. These software tools, resources, and databases drive the field forward to help keep pace with the amount of data being generated and the sophisticated and diverse analytical platforms that are being used to generate these metabolomics datasets. To address challenges in data preprocessing, metabolite annotation, statistical interrogation, visualization, interpretation, and integration, the metabolomics and informatics research community comes up with hundreds of tools every year. The purpose of the present review is to provide a brief and useful summary of more than 95 metabolomics tools, software, and databases that were either developed or significantly improved during 2017–2018. We hope to see this review help readers, developers, and researchers to obtain informed access to these thorough lists of resources for further improvisation, implementation, and application in due course of time.

Output data structure



Chromtography

Mass spectrometry

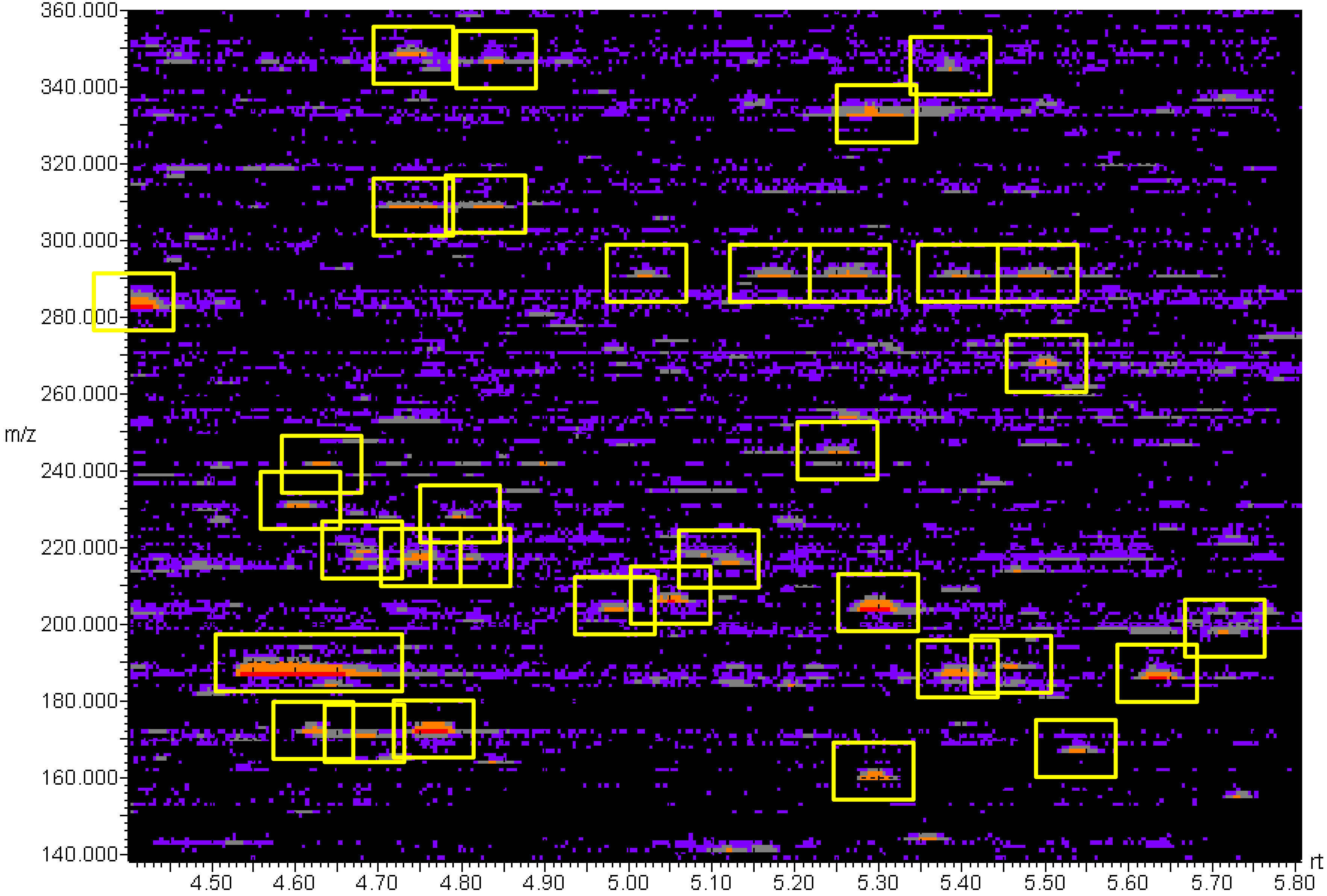
Detector

Retention time (RT)

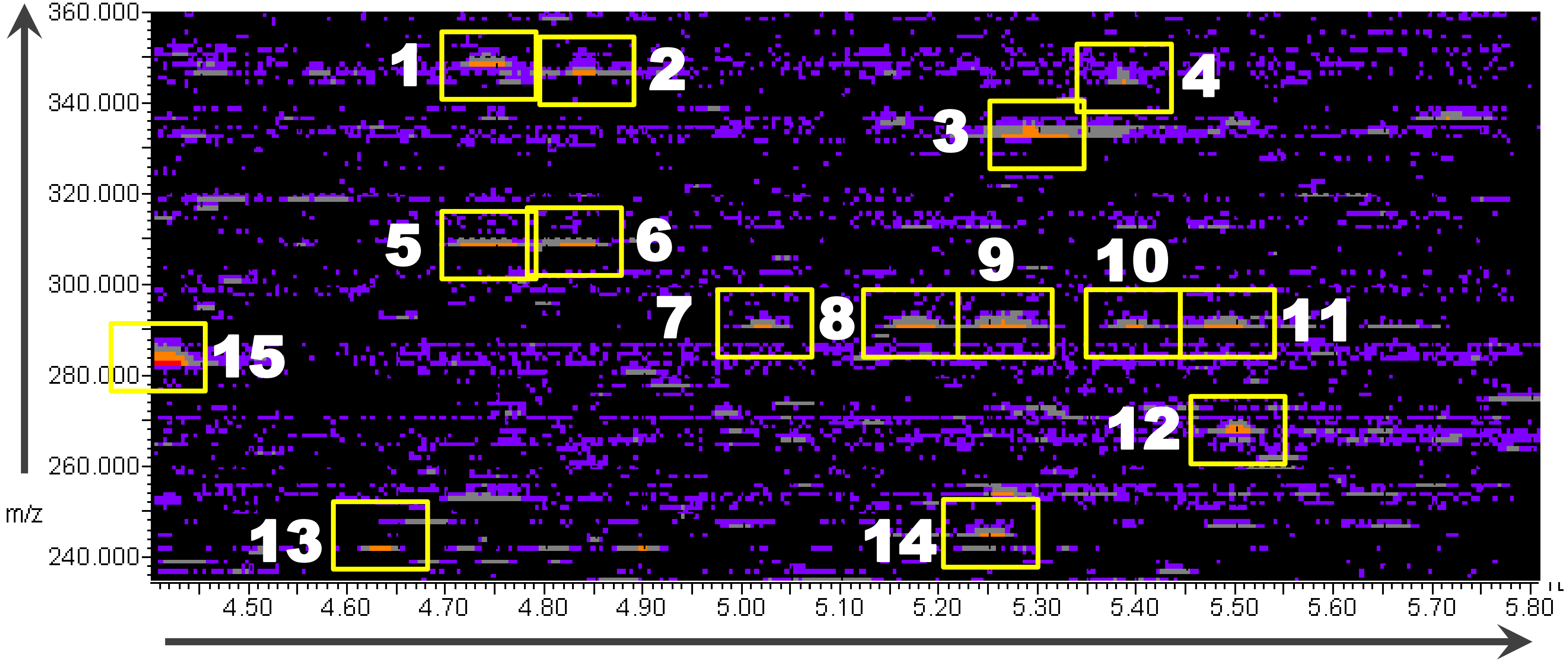
Mass to charge ratio (m/z)

Intensity

Feature detection in 3-dimensional data (LC-MS)



Feature detection in 3-dimensional data (LC-MS)



Feature detection in 3 dimensions

Concatenated to single term representing each feature

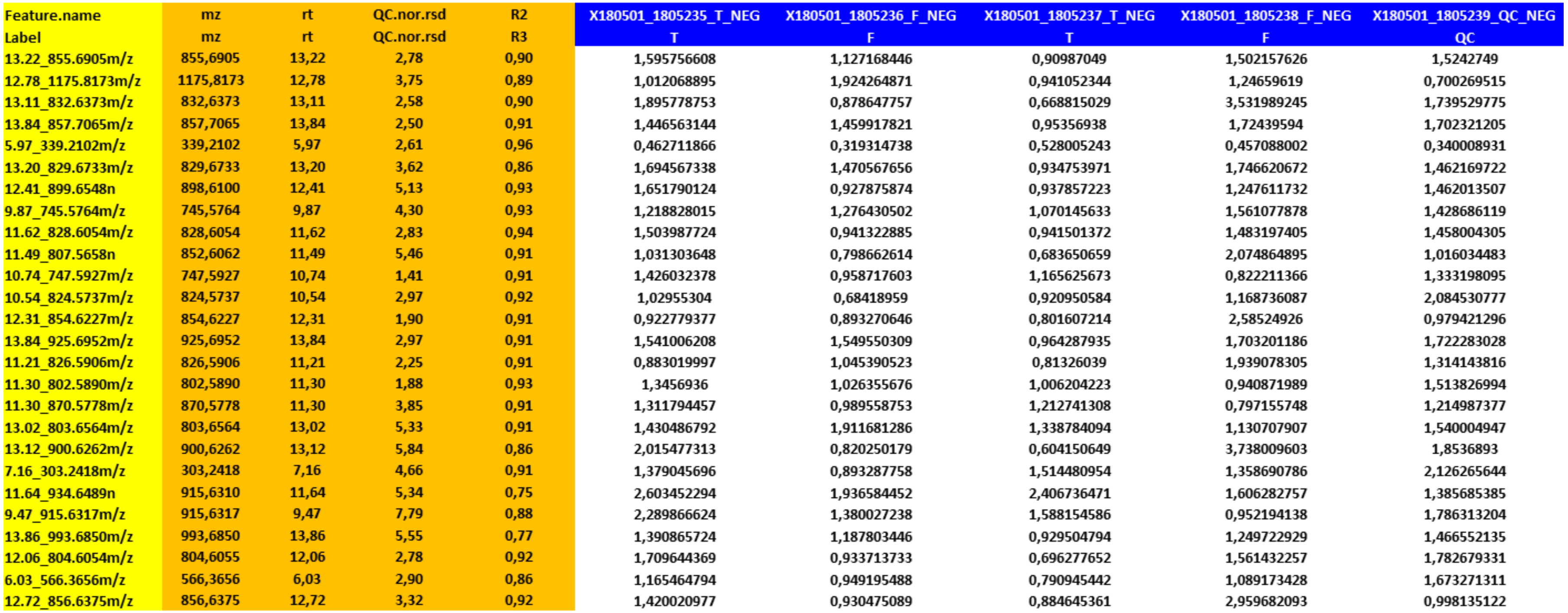
Feature = RT_mz

1. Mass (m/z)
2. Chromatographic retention time (RT)
3. Intensity (“counts”)

Output data structure

Feature Identifier

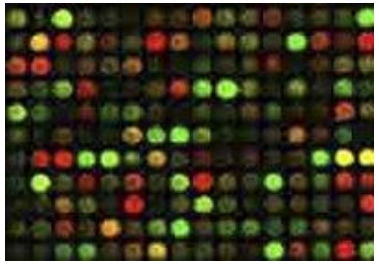
Sample identifier



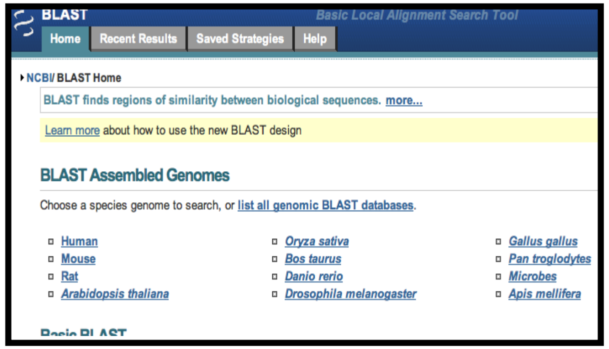
Feature.name	mz	rt	QC.nor.rsd	R2	X180501_1805235_T_NEG	X180501_1805236_F_NEG	X180501_1805237_T_NEG	X180501_1805238_F_NEG	X180501_1805239_QC_NEG
Label	mz	rt	QC.nor.rsd	R3	T	F	T	F	QC
13.22_855.6905m/z	855,6905	13,22	2,78	0,90	1,595756608	1,127168446	0,90987049	1,502157626	1,5242749
12.78_1175.8173m/z	1175,8173	12,78	3,75	0,89	1,012068895	1,924264871	0,941052344	1,24659619	0,700269515
13.11_832.6373m/z	832,6373	13,11	2,58	0,90	1,895778753	0,878647757	0,668815029	3,531989245	1,739529775
13.84_857.7065m/z	857,7065	13,84	2,50	0,91	1,446563144	1,459917821	0,95356938	1,72439594	1,702321205
5.97_339.2102m/z	339,2102	5,97	2,61	0,96	0,462711866	0,319314738	0,528005243	0,457088002	0,340008931
13.20_829.6733m/z	829,6733	13,20	3,62	0,86	1,694567338	1,470567656	0,934753971	1,746620672	1,462169722
12.41_899.6548n	898,6100	12,41	5,13	0,93	1,651790124	0,927875874	0,937857223	1,247611732	1,462013507
9.87_745.5764m/z	745,5764	9,87	4,30	0,93	1,218828015	1,276430502	1,070145633	1,561077878	1,428686119
11.62_828.6054m/z	828,6054	11,62	2,83	0,94	1,503987724	0,941322885	0,941501372	1,483197405	1,458004305
11.49_807.5658n	852,6062	11,49	5,46	0,91	1,031303648	0,798662614	0,683650659	2,074864895	1,016034483
10.74_747.5927m/z	747,5927	10,74	1,41	0,91	1,426032378	0,958717603	1,165625673	0,822211366	1,333198095
10.54_824.5737m/z	824,5737	10,54	2,97	0,92	1,02955304	0,68418959	0,920950584	1,168736087	2,084530777
12.31_854.6227m/z	854,6227	12,31	1,90	0,91	0,922779377	0,893270646	0,801607214	2,58524926	0,979421296
13.84_925.6952m/z	925,6952	13,84	2,97	0,91	1,541006208	1,549550309	0,964287935	1,703201186	1,722283028
11.21_826.5906m/z	826,5906	11,21	2,25	0,91	0,883019997	1,045390523	0,81326039	1,939078305	1,314143816
11.30_802.5890m/z	802,5890	11,30	1,88	0,93	1,3456936	1,026355676	1,006204223	0,940871989	1,513826994
11.30_870.5778m/z	870,5778	11,30	3,85	0,91	1,311794457	0,989558753	1,212741308	0,797155748	1,214987377
13.02_803.6564m/z	803,6564	13,02	5,33	0,91	1,430486792	1,911681286	1,338784094	1,130707907	1,540004947
13.12_900.6262m/z	900,6262	13,12	5,84	0,86	2,015477313	0,820250179	0,604150649	3,738009603	1,8536893
7.16_303.2418m/z	303,2418	7,16	4,66	0,91	1,379045696	0,893287758	1,514480954	1,358690786	2,126265644
11.64_934.6489n	915,6310	11,64	5,34	0,75	2,603452294	1,936584452	2,406736471	1,606282757	1,385685385
9.47_915.6317m/z	915,6317	9,47	7,79	0,88	2,289866624	1,380027238	1,588154586	0,952194138	1,786313204
13.86_993.6850m/z	993,6850	13,86	5,55	0,77	1,390865724	1,187803446	0,929504794	1,249722929	1,466552135
12.06_804.6054m/z	804,6055	12,06	2,78	0,92	1,709644369	0,933713733	0,696277652	1,561432257	1,782679331
6.03_566.3656m/z	566,3656	6,03	2,90	0,86	1,165464794	0,949195488	0,790945442	1,089173428	1,673271311
12.72_856.6375m/z	856,6375	12,72	3,32	0,92	1,420020977	0,930475089	0,884645361	2,959682093	0,998135122

Feature intensity

The biggest challenge is Annotation



DNA/RNA



BLAST Basic Local Alignment Search Tool

Home | Recent Results | Saved Strategies | Help

NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. more...

Learn more about how to use the new BLAST design

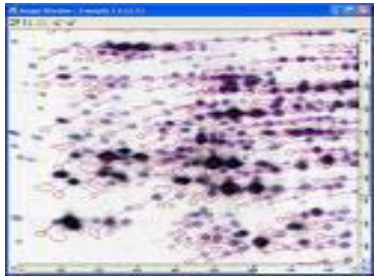
BLAST Assembled Genomes

Choose a species genome to search, or list all genomic BLAST databases.

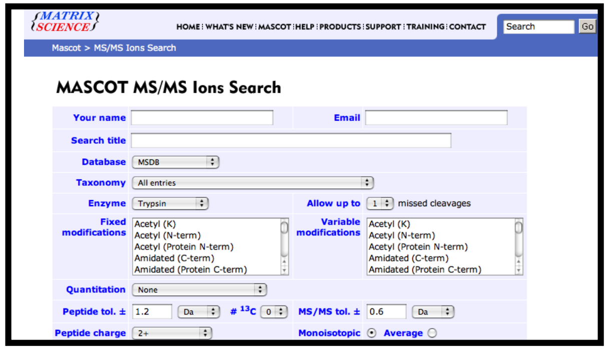
◻ Human	◻ Oryza sativa	◻ Gallus gallus
◻ Mouse	◻ Bos taurus	◻ Pan troglodytes
◻ Rat	◻ Danio rerio	◻ Microbes
◻ Arabidopsis thaliana	◻ Drosophila melanogaster	◻ Apis mellifera

Basic BLAST

Gene IDs +
Transcript
Abundance



Proteomics



MATRIX SCIENCE

HOME | WHAT'S NEW | MASCOT | HELP | PRODUCTS | SUPPORT | TRAINING | CONTACT   Search Go

Mascot > MS/MS Ions Search

MASCOT MS/MS Ions Search

Your name   Email

Search title

Database MSDB

Taxonomy All entries

Enzyme Trypsin   Allow up to 1 missed cleavages

Fixed modifications: Acetyl (K), Acetyl (N-term), Acetyl (Protein N-term), Amidated (C-term), Amidated (Protein C-term)

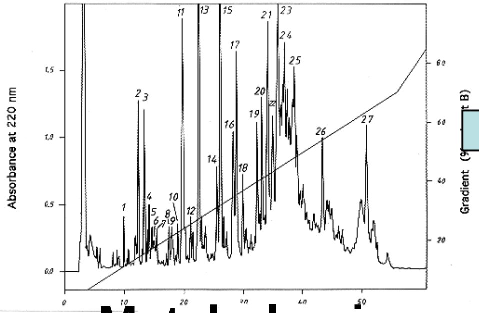
Variable modifications: Acetyl (K), Acetyl (N-term), Acetyl (Protein N-term), Amidated (C-term), Amidated (Protein C-term)

Quantitation None

Peptide tol. ± 1.2 Da   # ^{13}C 0   MS/MS tol. ± 0.6 Da

Peptide charge 2+   Monoisotopic ● Average ○

Protein IDs +
Concentrations



Metabolomics

?

Metabolite IDs +
Concentrations

Metabolome Databases

http://metabolomicssociety.org/resources/metabolomics-databases

www.hmdb.ca

www.drugbank.ca

www.ymdb.ca

www.phenol-explorer.eu

www.ecmdb.ca

www.foodb.ca

www.cowmetdb.ca

www.t3db.ca

www.smpdb.ca

www.csfmetabolome.ca

www.serummetabolome.ca

www.urinemetabolome.ca

The biggest challenge is Annotation

Levels of Metabolite Identification in MS

1. **Positively identified compounds**

   Confirmed by match to known standard

2. **Putatively identified compounds**

   Match to MS + RT or MS/MS + RT

3. **Compounds putatively identified in a compound class**
4. **Unknown compounds**

Processing tools

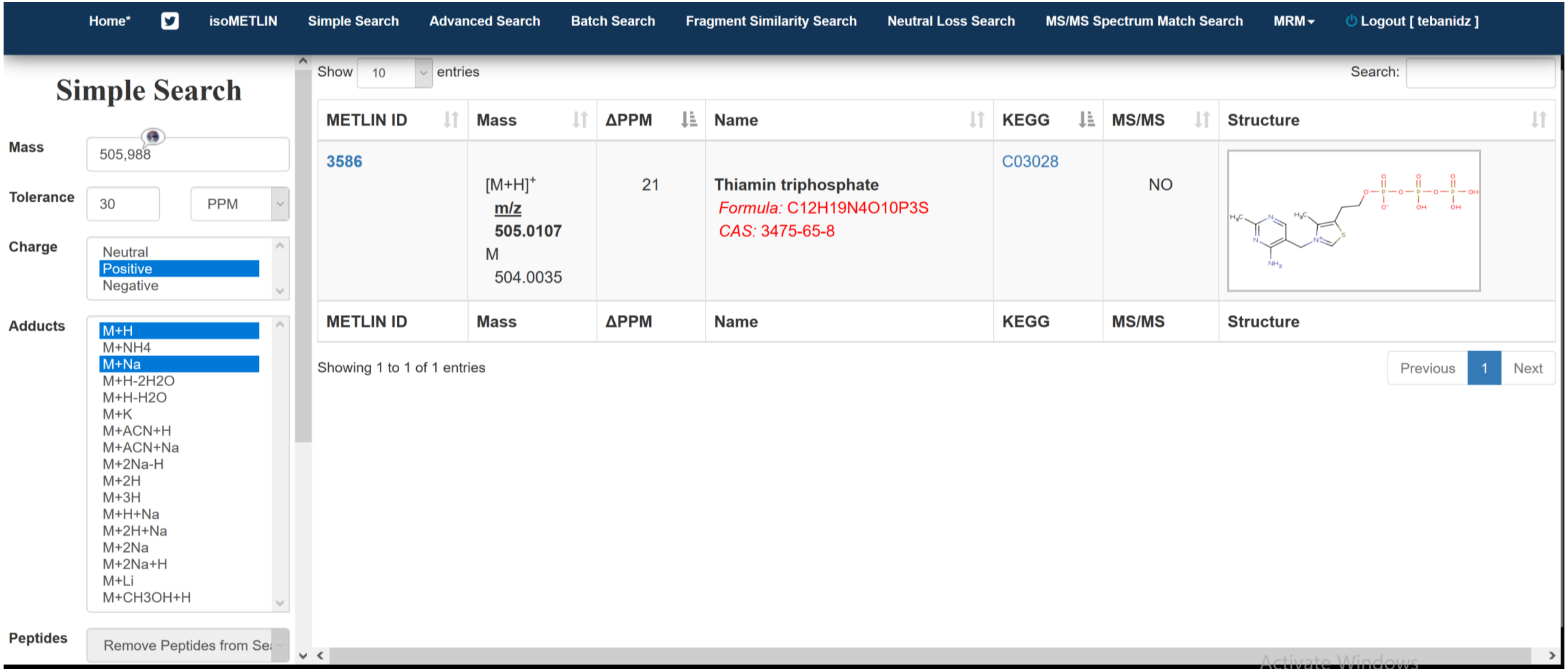
- Commercial tools
  - Agilent MassHunter Profinder
  - Bruker's ProfileAnalysis
  - Thermo SIEVE™
  - Waters' Progenesis QI
  - SkyLine
- Free options
  - XCMS Online
  - MZmine

Processing tools

XCMS

- The first open source tool for spectra processing
- Does peak picking, peak matching and retention time alignment
- Available as a program and a server
- Accepts multiple formats: mzXML, mzData, .cdf (NetCDF), .d folders (Agilent; Bruker), .wiff files (AB SCIEX)
- Metabolite identification is not the focus in XCMS (linked to Metlin)

METLIN



Simple Search

Mass: 505,988

Tolerance: 30 PPM

Charge: Neutral / **Positive** / Negative

Adducts: **M+H**, M+NH4, **M+Na**, M+H-2H2O, M+H-H2O, M+K, M+ACN+H, M+ACN+Na, M+2Na-H, M+2H, M+3H, M+H+Na, M+2H+Na, M+2Na, M+2Na+H, M+Li, M+CH3OH+H

Peptides: Remove Peptides from Se

Show 10 entries

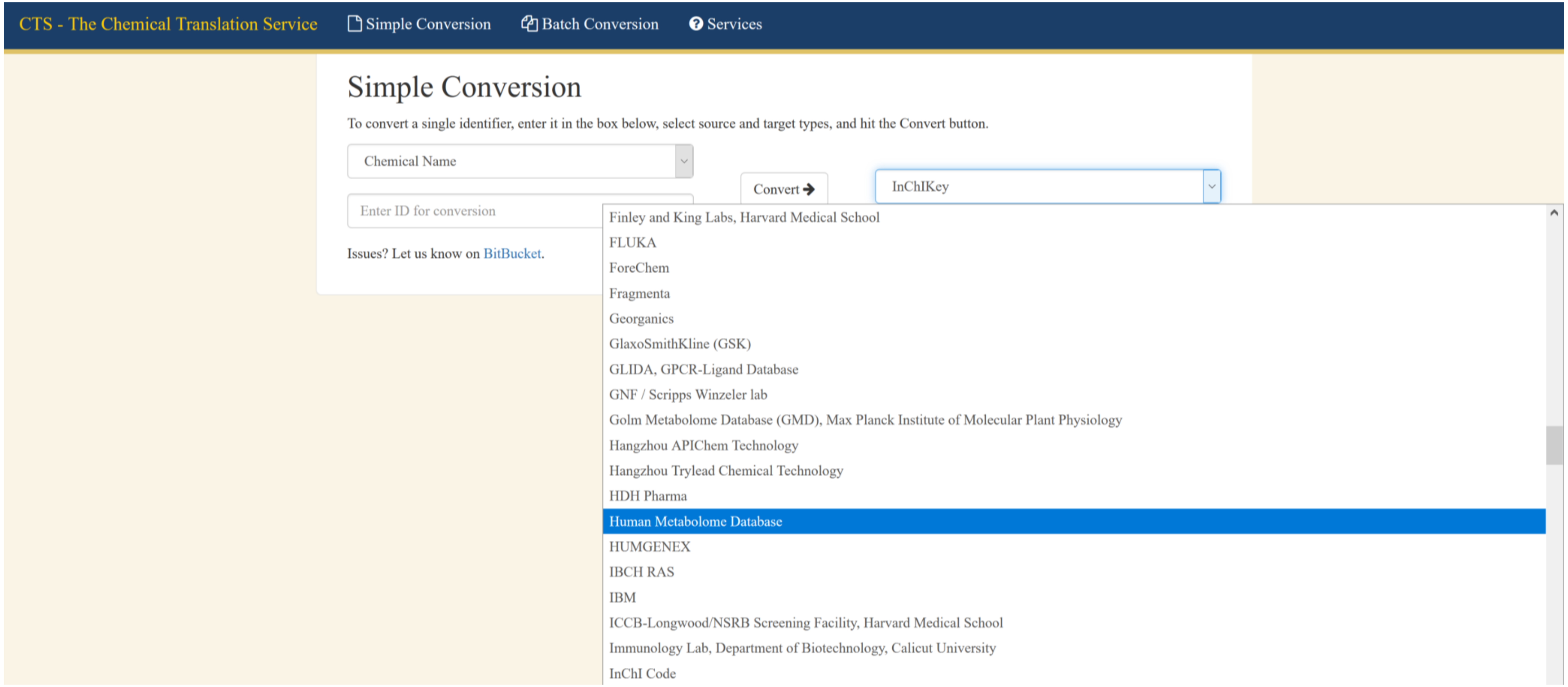
Search:

METLIN ID	Mass	ΔPPM	Name	KEGG	MS/MS	Structure
3586	$[M+H]^+$ **m/z** **505.0107** M 504.0035	21	**Thiamin triphosphate** *Formula:* C12H19N4O10P3S *CAS:* 3475-65-8	C03028	NO	
METLIN ID	Mass	ΔPPM	Name	KEGG	MS/MS	Structure

Showing 1 to 1 of 1 entries

Previous 1 Next

Annotation Conversion



CTS - The Chemical Translation Service

Simple Conversion

To convert a single identifier, enter it in the box below, select source and target types, and hit the Convert button.

Chemical Name

Enter ID for conversion

Convert ➔

InChIKey

Issues? Let us know on BitBucket.

- Finley and King Labs, Harvard Medical School
- FLUKA
- ForeChem
- Fragmenta
- Georganics
- GlaxoSmithKline (GSK)
- GLIDA, GPCR-Ligand Database
- GNF / Scripps Winzeler lab
- Golm Metabolome Database (GMD), Max Planck Institute of Molecular Plant Physiology
- Hangzhou APIChem Technology
- Hangzhou Trylead Chemical Technology
- HDH Pharma
- Human Metabolome Database
- HUMGENEX
- IBCH RAS
- IBM
- ICCB-Longwood/NSRB Screening Facility, Harvard Medical School
- Immunology Lab, Department of Biotechnology, Calicut University
- InChI Code

GETTING DATA READY

Data Analysis

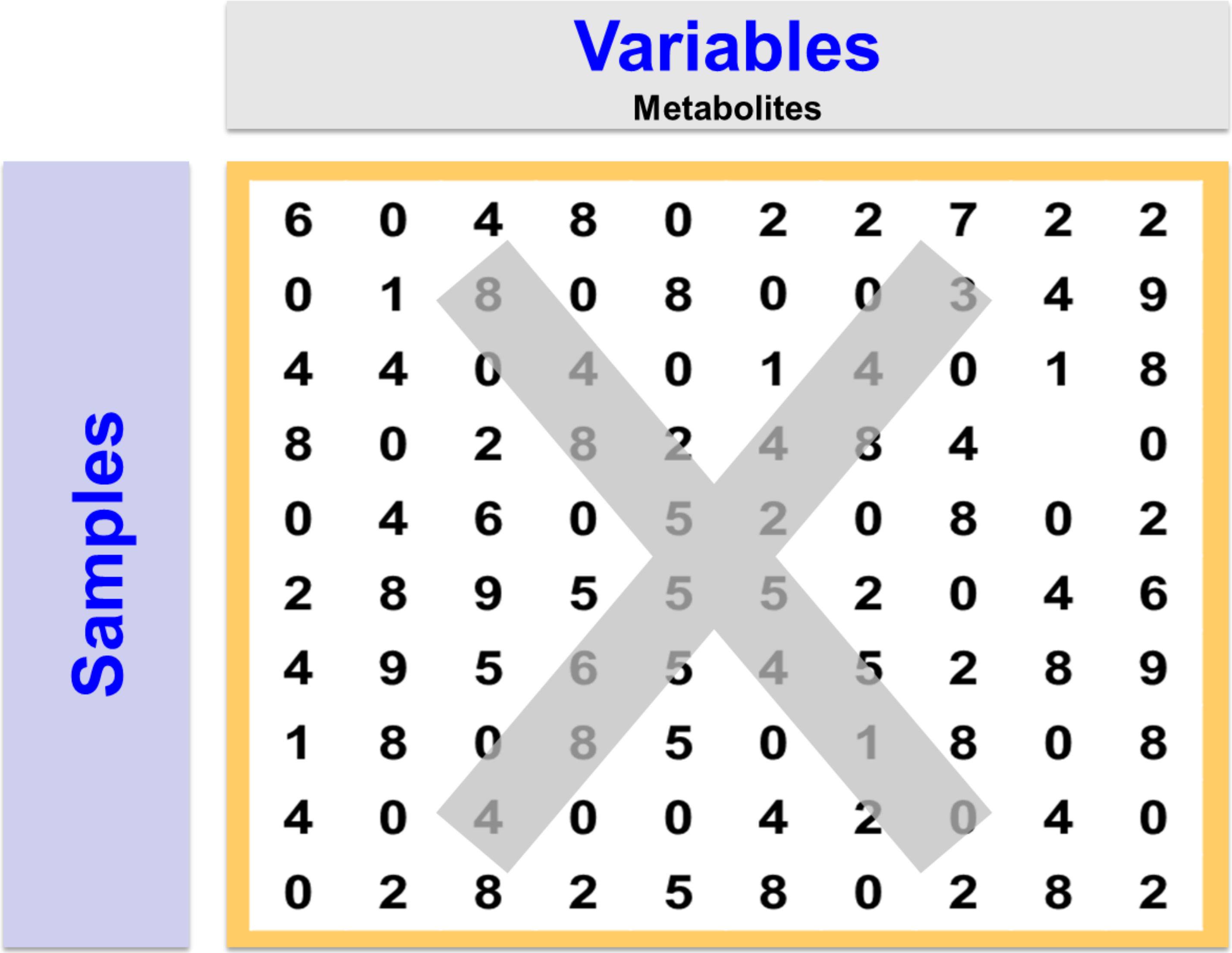
Input

- A matrix containing numerical values
    - Concentrations (Targeted)
    - Peak intensities (Untargeted)
- Meta-data
    - Class labels, experimental factors

Output

- Discriminant features
- Clustering patterns
- Biological Inference
- Biomarkers
- Predictive models

Data Analysis



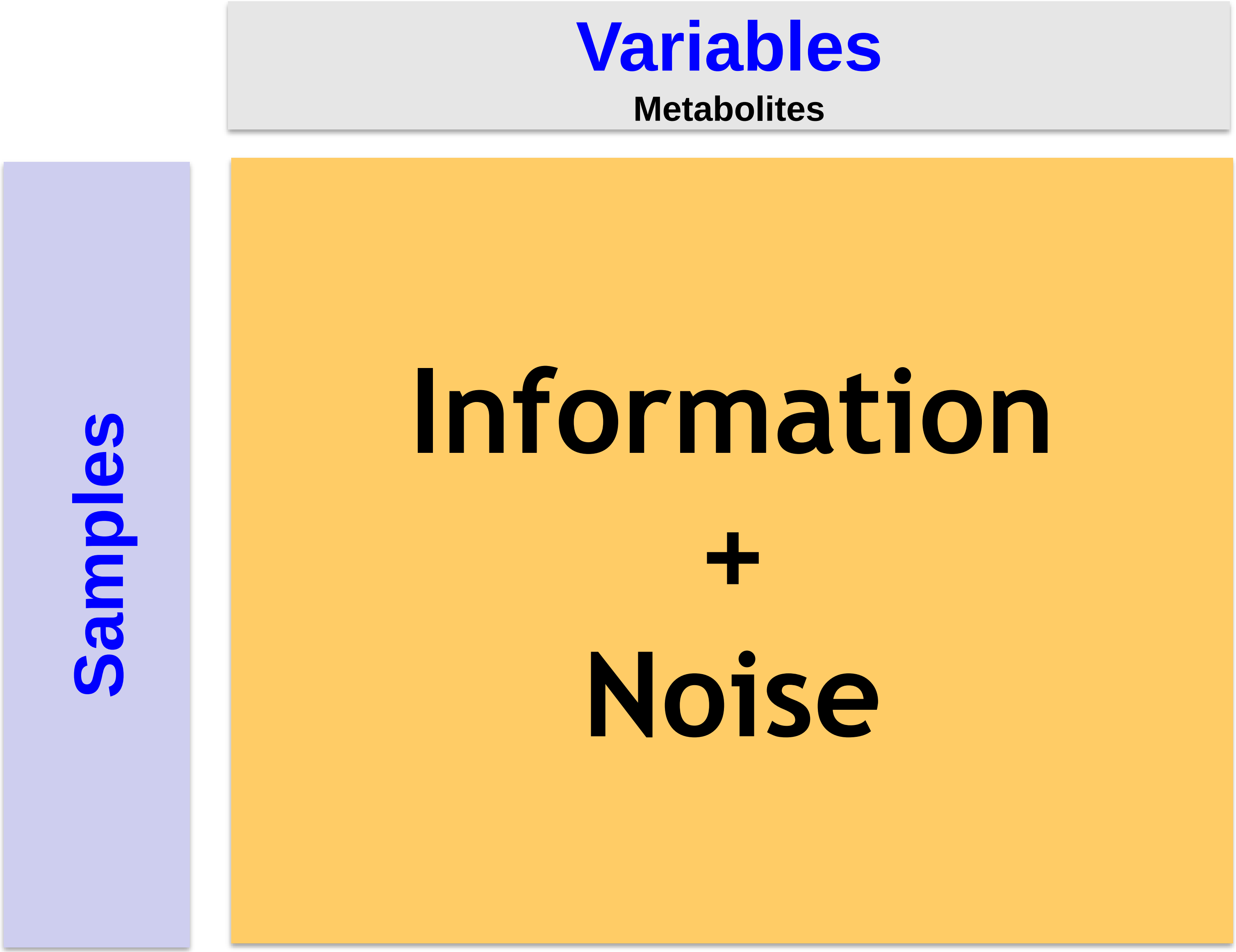
Variables

Metabolites

Samples

6	0	4	8	0	2	2	7	2	2
0	1	8	0	8	0	0	3	4	9
4	4	0	4	0	1	4	0	1	8
8	0	2	8	2	4	8	4		0
0	4	6	0	5	2	0	8	0	2
2	8	9	5	5	5	2	0	4	6
4	9	5	6	5	4	5	2	8	9
1	8	0	8	5	0	1	8	0	8
4	0	4	0	0	4	2	0	4	0
0	2	8	2	5	8	0	2	8	2

Data Analysis



Variables

Metabolites

Samples

Information

+

Noise

Data Analysis / Data cleaning

Remove as much as possible noise

Extract as much as possible information

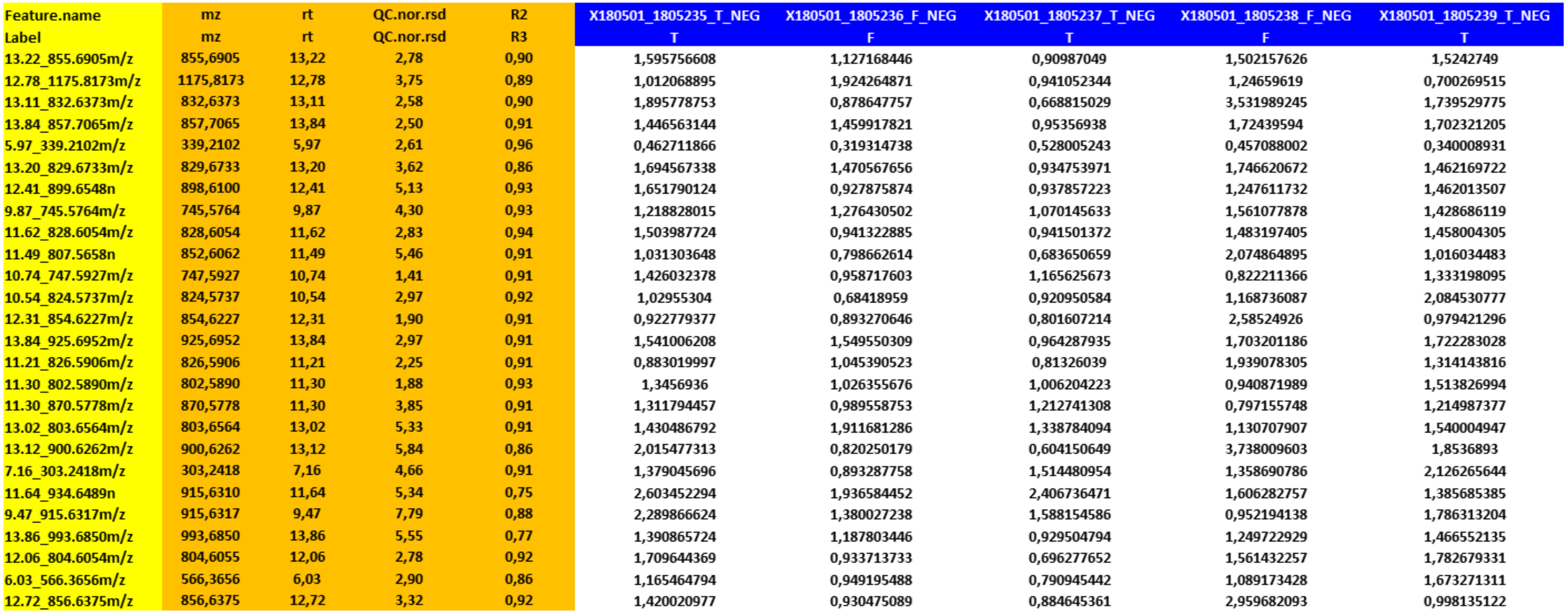
Data Analysis / Data cleaning

Data cleaning

- Missing values imputation
- Filtering (Min, IQR, RSD, CV, R2 …)
- Normalization

Data Analysis / Data cleaning

Cut-off



Feature.name	mz	rt	QC.nor.rsd	R2	X180501_1805235_T_NEG	X180501_1805236_F_NEG	X180501_1805237_T_NEG	X180501_1805238_F_NEG	X180501_1805239_T_NEG
Label	mz	rt	QC.nor.rsd	R3	T	F	T	F	T
13.22_855.6905m/z	855,6905	13,22	2,78	0,90	1,595756608	1,127168446	0,90987049	1,502157626	1,5242749
12.78_1175.8173m/z	1175,8173	12,78	3,75	0,89	1,012068895	1,924264871	0,941052344	1,24659619	0,700269515
13.11_832.6373m/z	832,6373	13,11	2,58	0,90	1,895778753	0,878647757	0,668815029	3,531989245	1,739529775
13.84_857.7065m/z	857,7065	13,84	2,50	0,91	1,446563144	1,459917821	0,95356938	1,72439594	1,702321205
5.97_339.2102m/z	339,2102	5,97	2,61	0,96	0,462711866	0,319314738	0,528005243	0,457088002	0,340008931
13.20_829.6733m/z	829,6733	13,20	3,62	0,86	1,694567338	1,470567656	0,934753971	1,746620672	1,462169722
12.41_899.6548n	898,6100	12,41	5,13	0,93	1,651790124	0,927875874	0,937857223	1,247611732	1,462013507
9.87_745.5764m/z	745,5764	9,87	4,30	0,93	1,218828015	1,276430502	1,070145633	1,561077878	1,428686119
11.62_828.6054m/z	828,6054	11,62	2,83	0,94	1,503987724	0,941322885	0,941501372	1,483197405	1,458004305
11.49_807.5658n	852,6062	11,49	5,46	0,91	1,031303648	0,798662614	0,683650659	2,074864895	1,016034483
10.74_747.5927m/z	747,5927	10,74	1,41	0,91	1,426032378	0,958717603	1,165625673	0,822211366	1,333198095
10.54_824.5737m/z	824,5737	10,54	2,97	0,92	1,02955304	0,68418959	0,920950584	1,168736087	2,084530777
12.31_854.6227m/z	854,6227	12,31	1,90	0,91	0,922779377	0,893270646	0,801607214	2,58524926	0,979421296
13.84_925.6952m/z	925,6952	13,84	2,97	0,91	1,541006208	1,549550309	0,964287935	1,703201186	1,722283028
11.21_826.5906m/z	826,5906	11,21	2,25	0,91	0,883019997	1,045390523	0,81326039	1,939078305	1,314143816
11.30_802.5890m/z	802,5890	11,30	1,88	0,93	1,3456936	1,026355676	1,006204223	0,940871989	1,513826994
11.30_870.5778m/z	870,5778	11,30	3,85	0,91	1,311794457	0,989558753	1,212741308	0,797155748	1,214987377
13.02_803.6564m/z	803,6564	13,02	5,33	0,91	1,430486792	1,911681286	1,338784094	1,130707907	1,540004947
13.12_900.6262m/z	900,6262	13,12	5,84	0,86	2,015477313	0,820250179	0,604150649	3,738009603	1,8536893
7.16_303.2418m/z	303,2418	7,16	4,66	0,91	1,379045696	0,893287758	1,514480954	1,358690786	2,126265644
11.64_934.6489n	915,6310	11,64	5,34	0,75	2,603452294	1,936584452	2,406736471	1,606282757	1,385685385
9.47_915.6317m/z	915,6317	9,47	7,79	0,88	2,289866624	1,380027238	1,588154586	0,952194138	1,786313204
13.86_993.6850m/z	993,6850	13,86	5,55	0,77	1,390865724	1,187803446	0,929504794	1,249722929	1,466552135
12.06_804.6054m/z	804,6055	12,06	2,78	0,92	1,709644369	0,933713733	0,696277652	1,561432257	1,782679331
6.03_566.3656m/z	566,3656	6,03	2,90	0,86	1,165464794	0,949195488	0,790945442	1,089173428	1,673271311
12.72_856.6375m/z	856,6375	12,72	3,32	0,92	1,420020977	0,930475089	0,884645361	2,959682093	0,998135122

Data Analysis / Normalization

Sample normalization (row-wise)

To remove systematic variation between experimental conditions **unrelated** to the biological differences (i.e. dilutions, mass)

Total signal, sum of signals
Reference compound: Internal standards, endogenous metabolites
Reference sample: QCs, Controls…

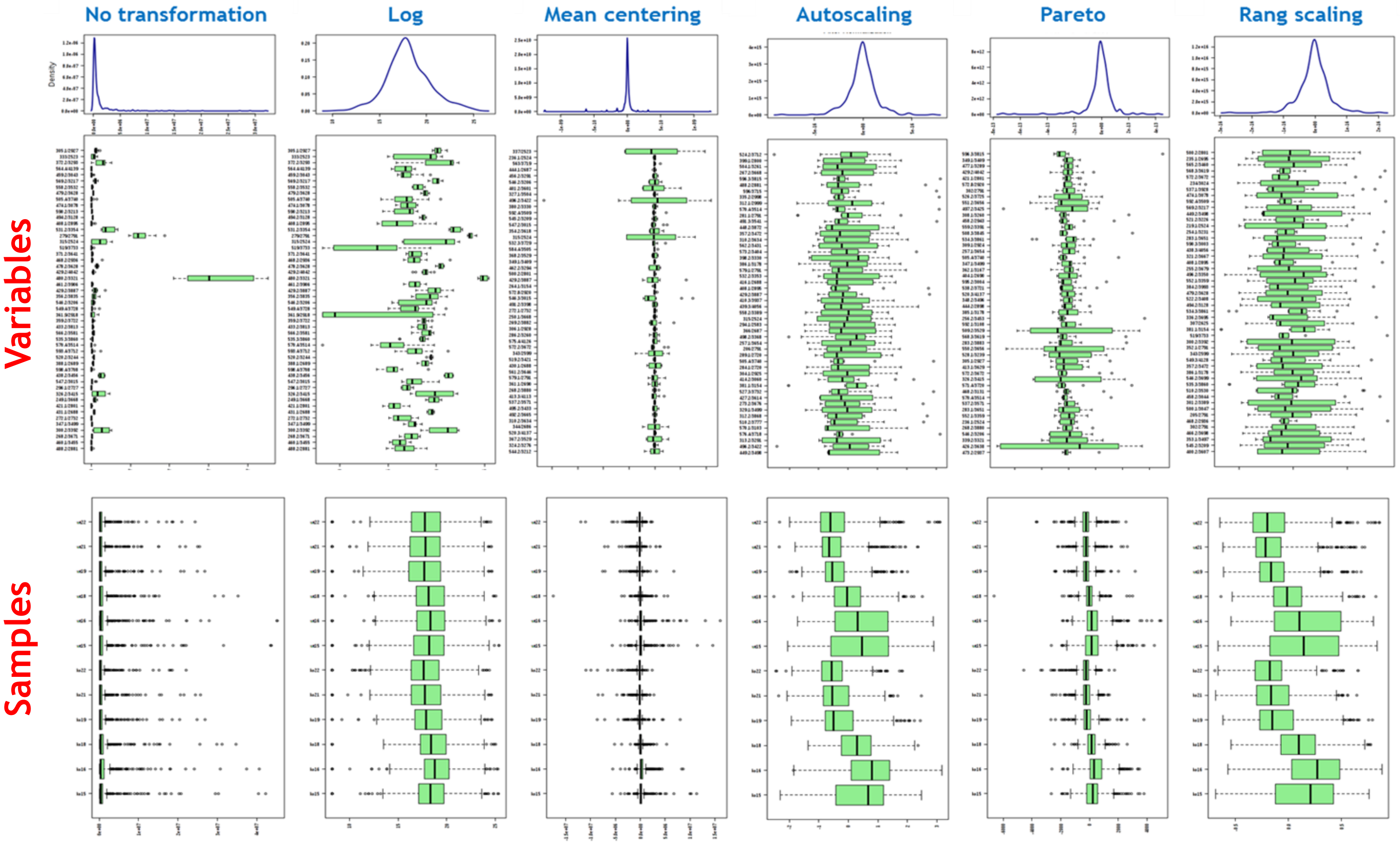
Feature normalization (column-wise)

To bring variances of all features close to equal

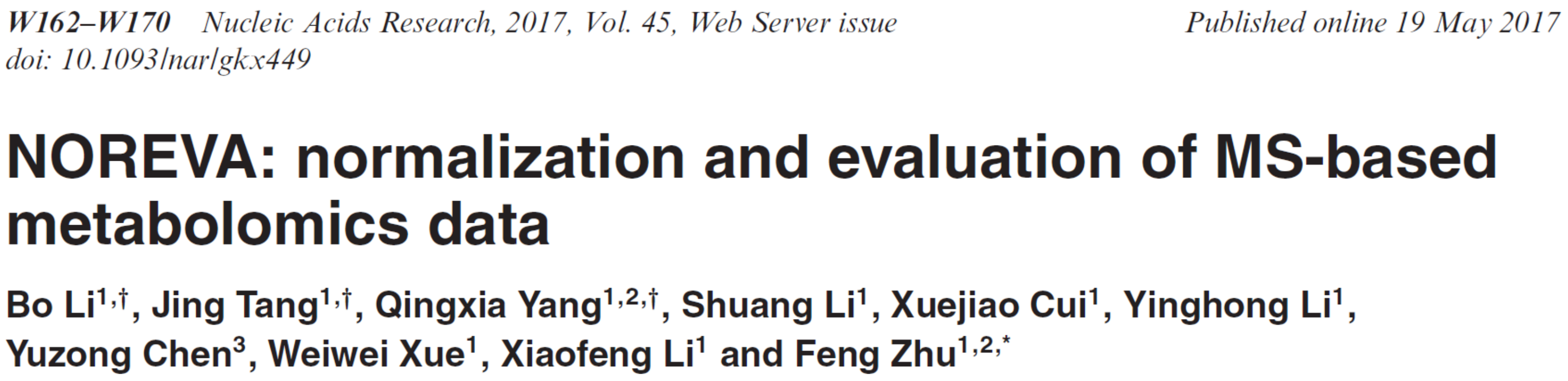
Log transformation

Scaling

Data Analysis / Normalization



Data Analysis / Normalization



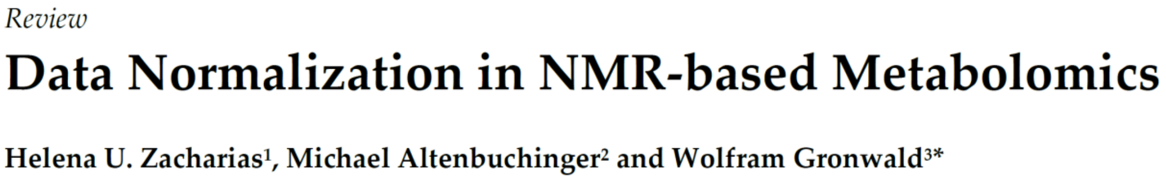
W162–W170 Nucleic Acids Research, 2017, Vol. 45, Web Server issue *Published online 19 May 2017*
doi: 10.1093/nar/gkx449

NOREVA: normalization and evaluation of MS-based metabolomics data

Bo Li[1,†], Jing Tang[1,†], Qingxia Yang[1,2,†], Shuang Li[1], Xuejiao Cui[1], Yinghong Li[1], Yuzong Chen[3], Weiwei Xue[1], Xiaofeng Li[1] and Feng Zhu[1,2,*]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 July 2018 doi:10.20944/preprints201807.0059.v1

Peer-reviewed version available at *Metabolites* **2018**, *8*, 47; doi:10.3390/metabo8030047



Review

Data Normalization in NMR-based Metabolomics

Helena U. Zacharias[1], Michael Altenbuchinger[2] and Wolfram Gronwald[3*]

Multivariate Data Analysis

Tow main objectives

Descriptive data analysis (Unsupervised learning)

Mining massive datasets to discover hidden

- *data structures*
- *hidden relationships*
- *patterns, trends and clusters*
- *outliers*

Dimension reduction

X

Predictive data analysis (Supervised learning)

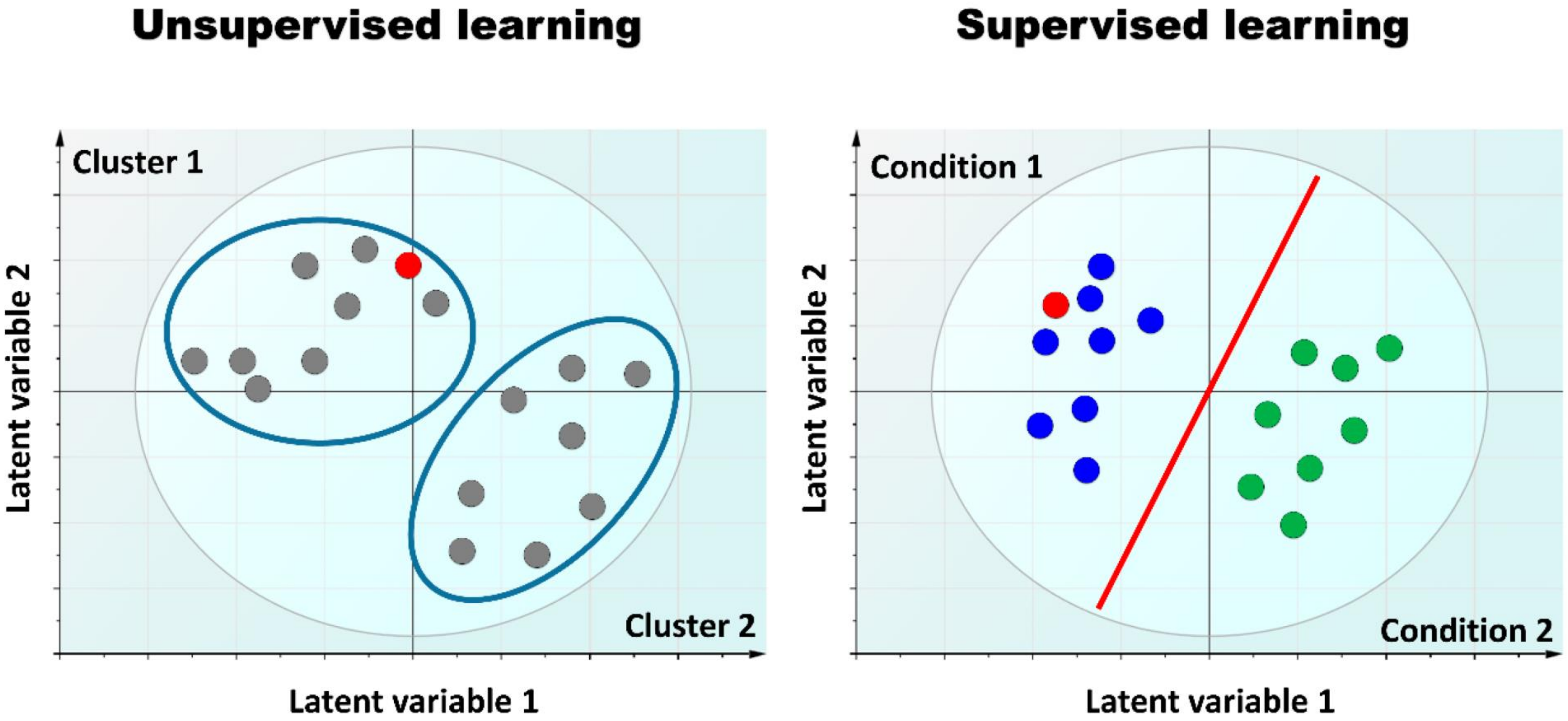
Building models for specific tasks using training datasets

- *regression*
- *classification,*
- *pattern recognition*
- *machine learning tasks*

Assessing the predictive accuracy of the models using new datasets

X Y

Multivariate Data Analysis



Tebani A *et.al*. Int J Mol Sci. 2016 Sep; 17(9): 1.

Data Analysis / Unsupervised learning

Clustering

Organize the 1000s of variables into blocks

Variables in each block are more homogenous

Key parameter: similarities (Distance, Spearman, Pearson …)

Similarity between samples - Similarity between clusters

Visualization using Heatmaps

- *K-means*
- *Hierarchical Methods*

Dimension reduction

Reduce the high-dimensional data

1000s into low-dimensions (Latent variables)

- *Principal component analysis (PCA)*

Data Analysis / Supervised learning

Linear Discriminant Analysis
Partial Least Squares
k-Nearest Neighbors
Random Forest
Support Vector Machines
Bayesian networks
Neural Networks

… name it

https://jawbone.com/blog/classifying-steps-machine-learning/

Model validation

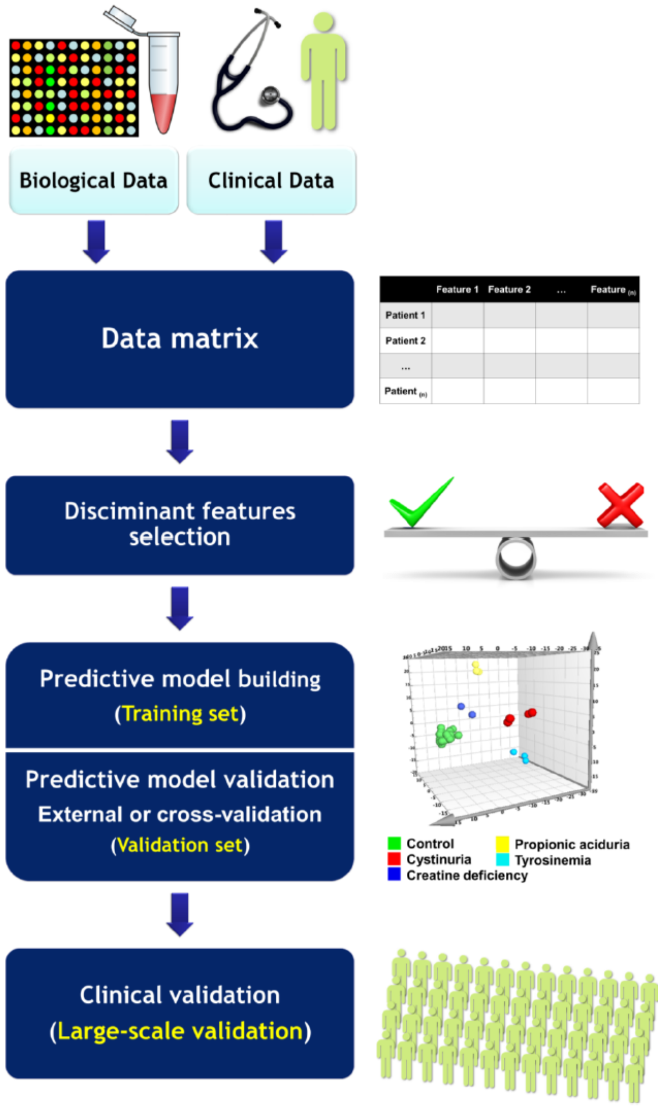
A *statistically valid* model

- Has good fit to the data
- Is *predictive* of new data

Validation methods

- Training set / Testing set
- Cross-validation
- Permutation test

Data Analysis / Supervised learning



Biological Data

Clinical Data

Data matrix

	Feature 1	Feature 2	...	Feature $_{(n)}$
Patient 1				
Patient 2				
...				
Patient $_{(n)}$				

Discriminant features selection

Predictive model building (Training set)

Predictive model validation External or cross-validation (Validation set)

Clinical validation (Large-scale validation)

Tebani *et al.* Int J Mol Sci. 2016 Sep; 17(9): 1.

Training dataset: a set of examples used to build the predictive model

Validation dataset: a set of examples used to refine the model parameters and estimate the error.

Test dataset: used only to evaluate the predictive performance of the model. They are never used during the learning or testing process.

Take home message

Experimental design +++++++

Know you data

Main technologies are MS and NMR

Annotation is challenging in untargeted metabolomics