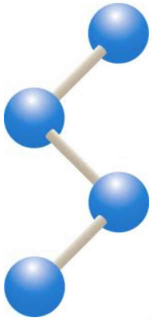
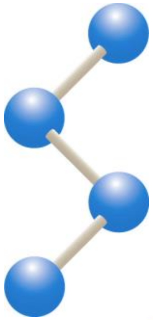
Fundamental & Advanced Metabolomics



May 16th-17th

Fundamental & Advanced Metabolomics



May 16th-17th

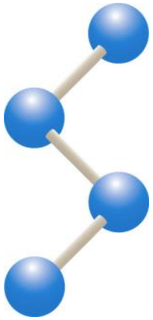
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

Fundamental & Advanced Metabolomics



May 16th-17th

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

May 17th

---- 09:00 am Begin ----

---- 10:30 am Break ----

---- 12:00 pm Lunch ---

---- 02:30 pm Break ----

---- 04:00 pm Finish ----

Fundamental & Advanced Metabolomics

Gary
Siuzdak

Fundamental & Advanced Metabolomics

Gary
Siuzdak

Elizabeth
Billings

Fundamental & Advanced Metabolomics

Gary
Siuzdak

Elizabeth
Billings

Amelia
Palermo

Fundamental & Advanced Metabolomics

Gary
Siuzdak

Elizabeth
Billings

Amelia
Palermo

Tao
Huan

Fundamental & Advanced Metabolomics


Gary
Siuzdak


Elizabeth
Billings


Amelia
Palermo


Tao
Huan


Bill
Webb

Fundamental & Advanced Metabolomics


Gary
Siuzdak


Elizabeth
Billings


Amelia
Palermo


Tao
Huan


Bill
Webb


H. Paul
Benton


Xavi
Domingo


Erica
Forsberg


Carlos
Guijas


Rafa
Montenegro

May 16th

Fundamental Metabolomics

- ***Objectives and Challenges***

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

May 16th

Fundamental Metabolomics

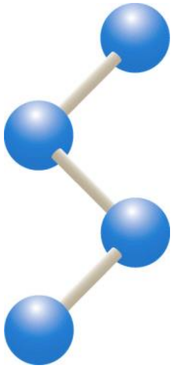
- ***Objectives and Challenges***
- ***Experimental Design***

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***

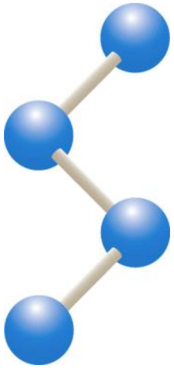
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***

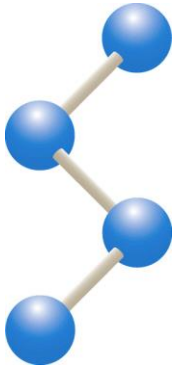
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***

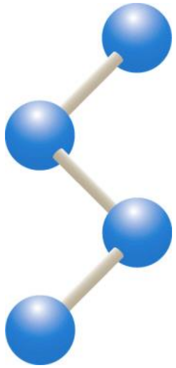
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

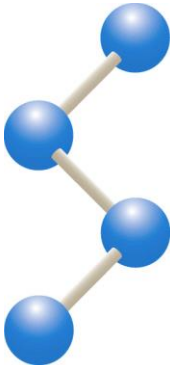
May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Statistical Analysis***

May 16th
---- 12:30 pm Begin ----
---- 3:00 pm Break ----
---- 04:30 pm Finish ----



Advanced Metabolomics

May 17th

---- 09:00 am Begin ----

---- 10:30 am Break ----

---- 12:00 pm Lunch ---

---- 02:30 pm Break ----

---- 04:00 pm Finish ----

Advanced Metabolomics

- ***Primary Experimental and Informatic Challenges***

May 17th

---- 09:00 am Begin ----

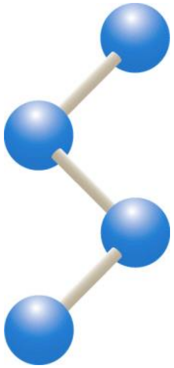
---- 10:30 am Break ----

---- 12:00 pm Lunch ---

---- 02:30 pm Break ----

---- 04:00 pm Finish ----

May 17th



Advanced Metabolomics

- ***Primary Experimental and Informatic Challenges***
- ***Key Algorithms in Creating Reproducible Data***

May 17th

---- 09:00 am Begin ----

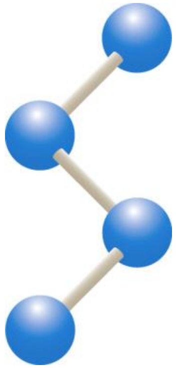
---- 10:30 am Break ----

---- 12:00 pm Lunch ---

---- 02:30 pm Break ----

---- 04:00 pm Finish ----

May 17th



Advanced Metabolomics

- ***Primary Experimental and Informatic Challenges***
- ***Key Algorithms in Creating Reproducible Data***
- ***Computational Metabolite Data Annotation***

May 17th

---- 09:00 am Begin ----

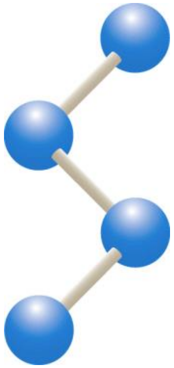
---- 10:30 am Break ----

---- 12:00 pm Lunch ---

---- 02:30 pm Break ----

---- 04:00 pm Finish ----

May 17th



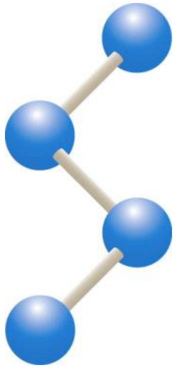
Advanced Metabolomics

- ***Primary Experimental and Informatic Challenges***
- ***Key Algorithms in Creating Reproducible Data***
- ***Computational Metabolite Data Annotation***
- ***Pathway Analysis & Multi-Omic Integration***

May 17th

---- 09:00 am Begin ----
---- 10:30 am Break ----
---- 12:00 pm Lunch ---
---- 02:30 pm Break ----
---- 04:00 pm Finish ----

May 17th

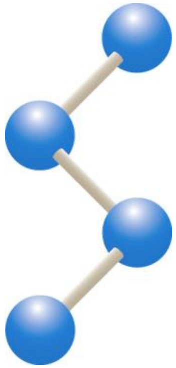


Advanced Metabolomics

- ***Primary Experimental and Informatic Challenges***
- ***Key Algorithms in Creating Reproducible Data***
- ***Computational Metabolite Data Annotation***
- ***Pathway Analysis & Multi-Omic Integration***
- ***Identifying Metabolites from Scratch***

May 17th
---- 09:00 am Begin ----
---- 10:30 am Break ----
---- 12:00 pm Lunch ---
---- 02:30 pm Break ----
---- 04:00 pm Finish ----

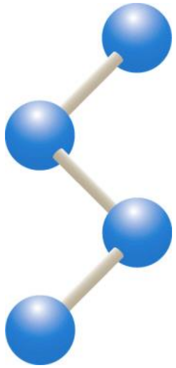
May 17th



Advanced Metabolomics

- ***Primary Experimental and Informatic Challenges***
- ***Key Algorithms in Creating Reproducible Data***
- ***Computational Metabolite Data Annotation***
- ***Pathway Analysis & Multi-Omic Integration***
- ***Identifying Metabolites from Scratch***
- ***Statistics in Design & Interpretation***

May 17th
---- 09:00 am Begin ----
---- 10:30 am Break ----
---- 12:00 pm Lunch ---
---- 02:30 pm Break ----
---- 04:00 pm Finish ----



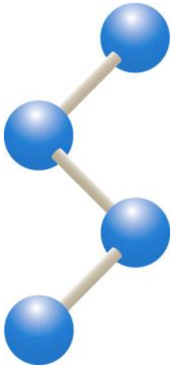
Advanced Metabolomics

- ***Primary Experimental and Informatic Challenges***
- ***Key Algorithms in Creating Reproducible Data***
- ***Computational Metabolite Data Annotation***
- ***Pathway Analysis & Multi-Omic Integration***
- ***Identifying Metabolites from Scratch***
- ***Statistics in Design & Interpretation***
- ***Activity Metabolomics***

May 17th

---- 09:00 am Begin ----
---- 10:30 am Break ----
---- 12:00 pm Lunch ---
---- 02:30 pm Break ----
---- 04:00 pm Finish ----

May 16th



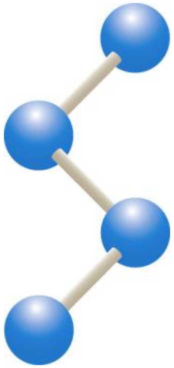
Fundamental Metabolomics

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***

May 16th

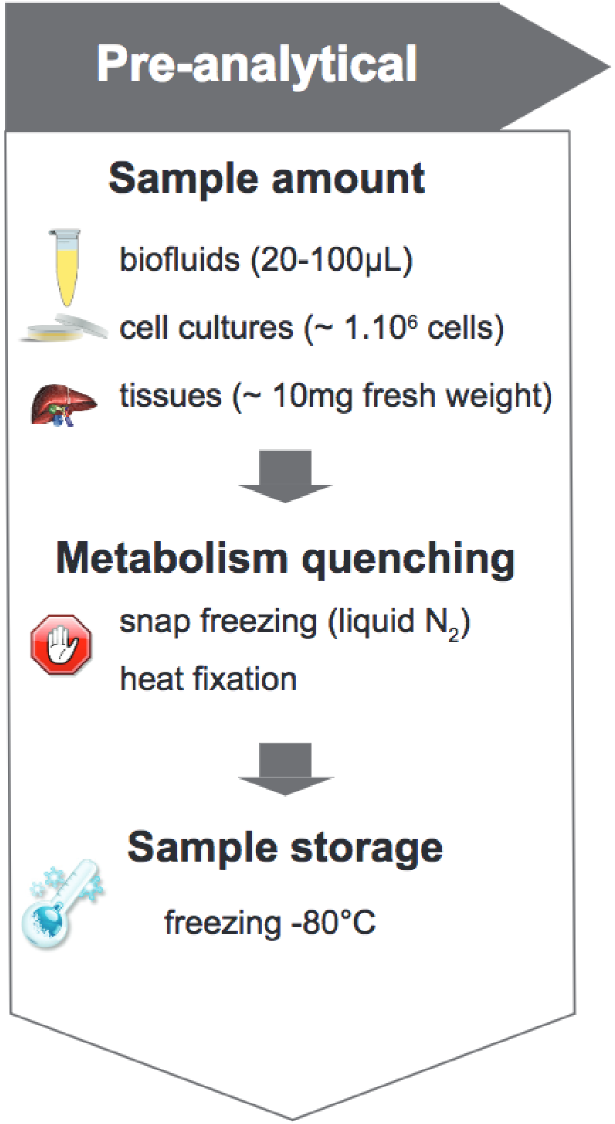
---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

Experimental Design

Fundamental Metabolomics



Pre-analytical

Sample amount

- biofluids (20-100µL)
- cell cultures (~ 1.10^6 cells)
- tissues (~ 10mg fresh weight)

Metabolism quenching

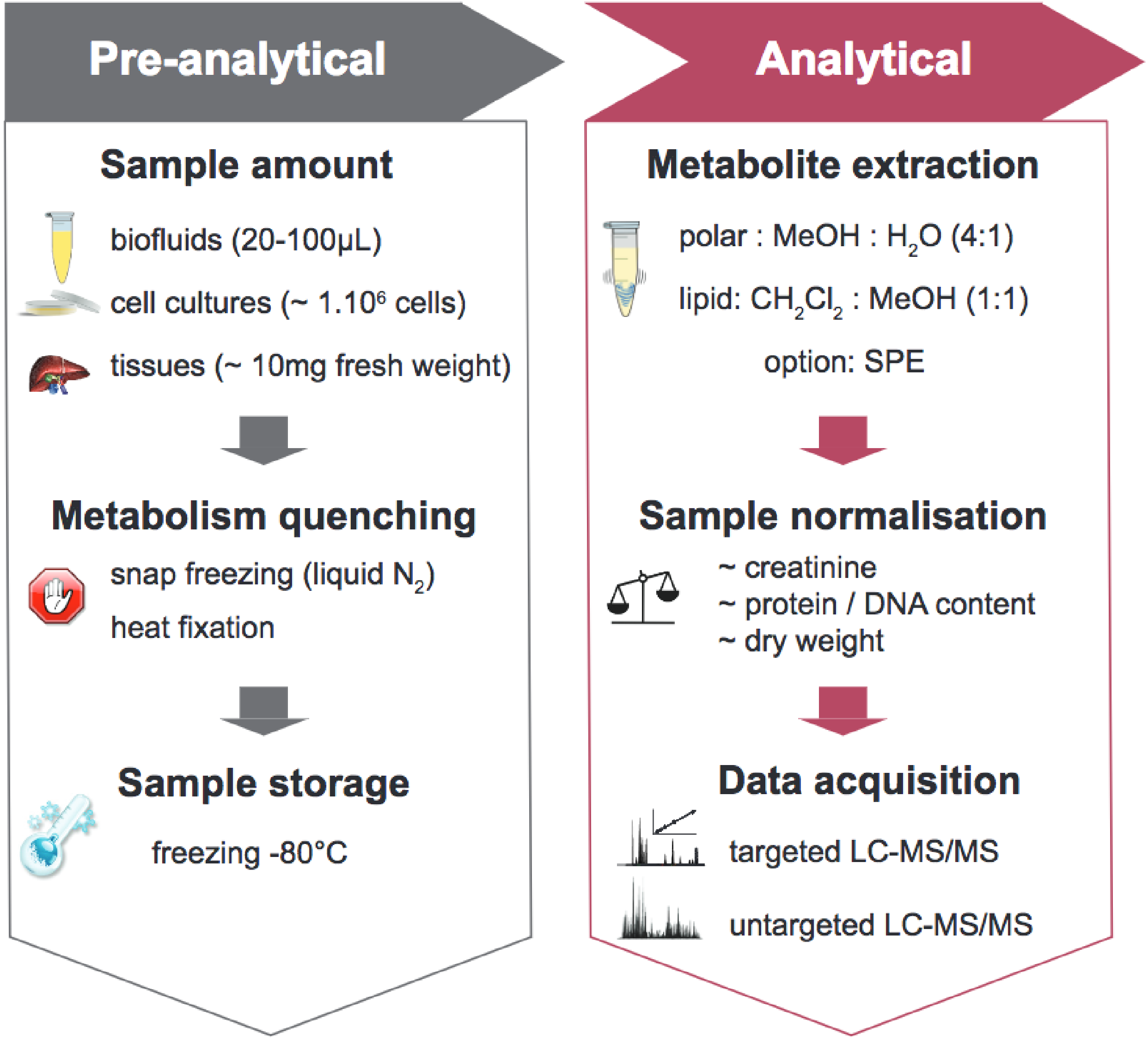
- snap freezing (liquid N_2)
- heat fixation

Sample storage

- freezing -80°C

Experimental Design

Fundamental Metabolomics



Pre-analytical

Sample amount

biofluids (20-100µL)

cell cultures (~ 1.10^6 cells)

tissues (~ 10mg fresh weight)

Metabolism quenching

snap freezing (liquid N_2)

heat fixation

Sample storage

freezing -80°C

Analytical

Metabolite extraction

polar : MeOH : H_2O (4:1)

lipid: CH_2Cl_2 : MeOH (1:1)

option: SPE

Sample normalisation

~ creatinine

~ protein / DNA content

~ dry weight

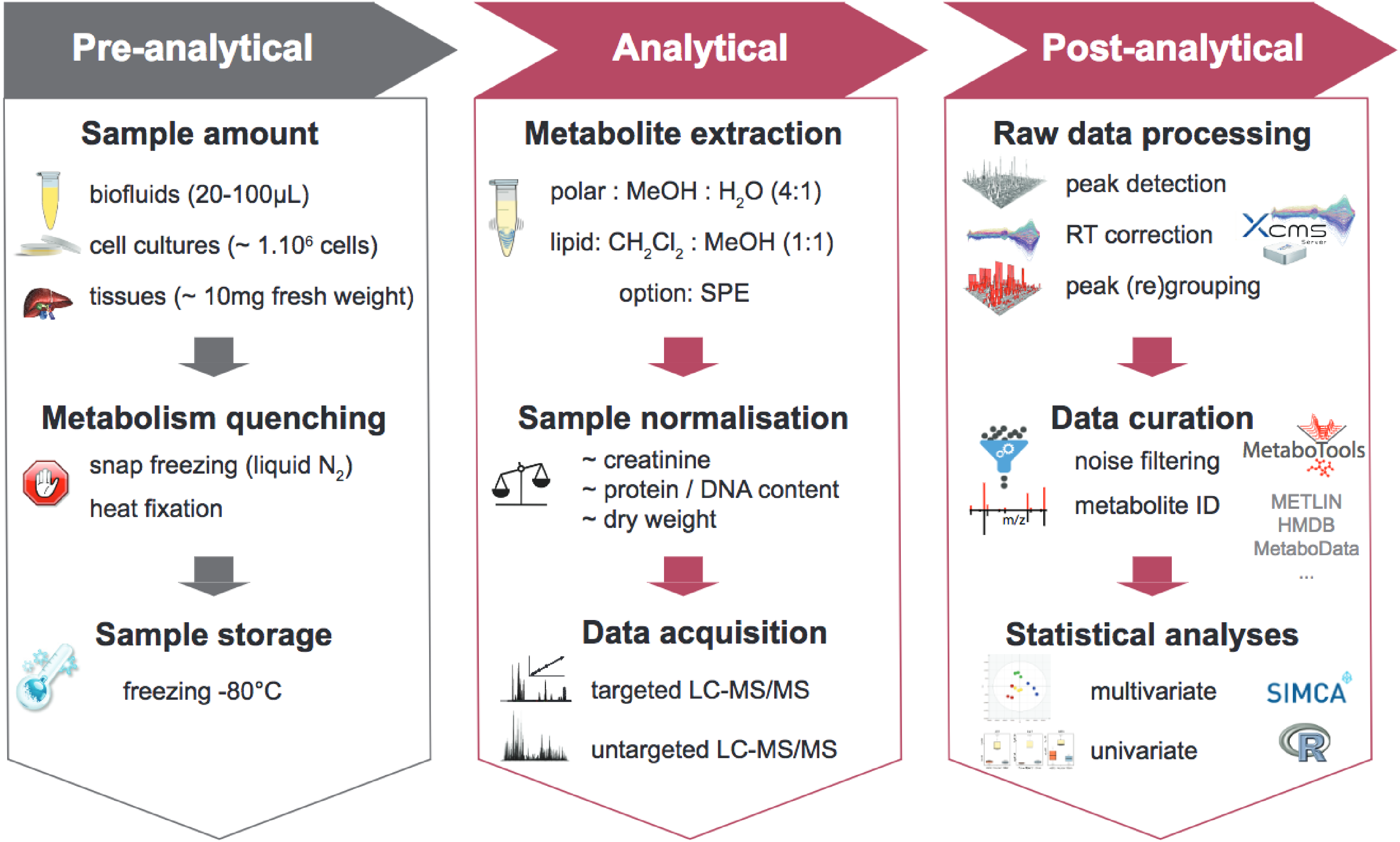
Data acquisition

targeted LC-MS/MS

untargeted LC-MS/MS

Experimental Design

Fundamental Metabolomics Advanced



Pre-analytical

Sample amount

- biofluids (20-100µL)
- cell cultures (~ 1.10⁶ cells)
- tissues (~ 10mg fresh weight)

Metabolism quenching

- snap freezing (liquid N_2)
- heat fixation

Sample storage

- freezing -80°C

Analytical

Metabolite extraction

- polar : MeOH : H_2O (4:1)
- lipid: CH_2Cl_2 : MeOH (1:1)
- option: SPE

Sample normalisation

- ~ creatinine
- ~ protein / DNA content
- ~ dry weight

Data acquisition

- targeted LC-MS/MS
- untargeted LC-MS/MS

Post-analytical

Raw data processing

- peak detection
- RT correction
- peak (re)grouping

XCMS Server

Data curation

- noise filtering
- metabolite ID

MetaboTools

METLIN
HMDB
MetaboData
...

Statistical analyses

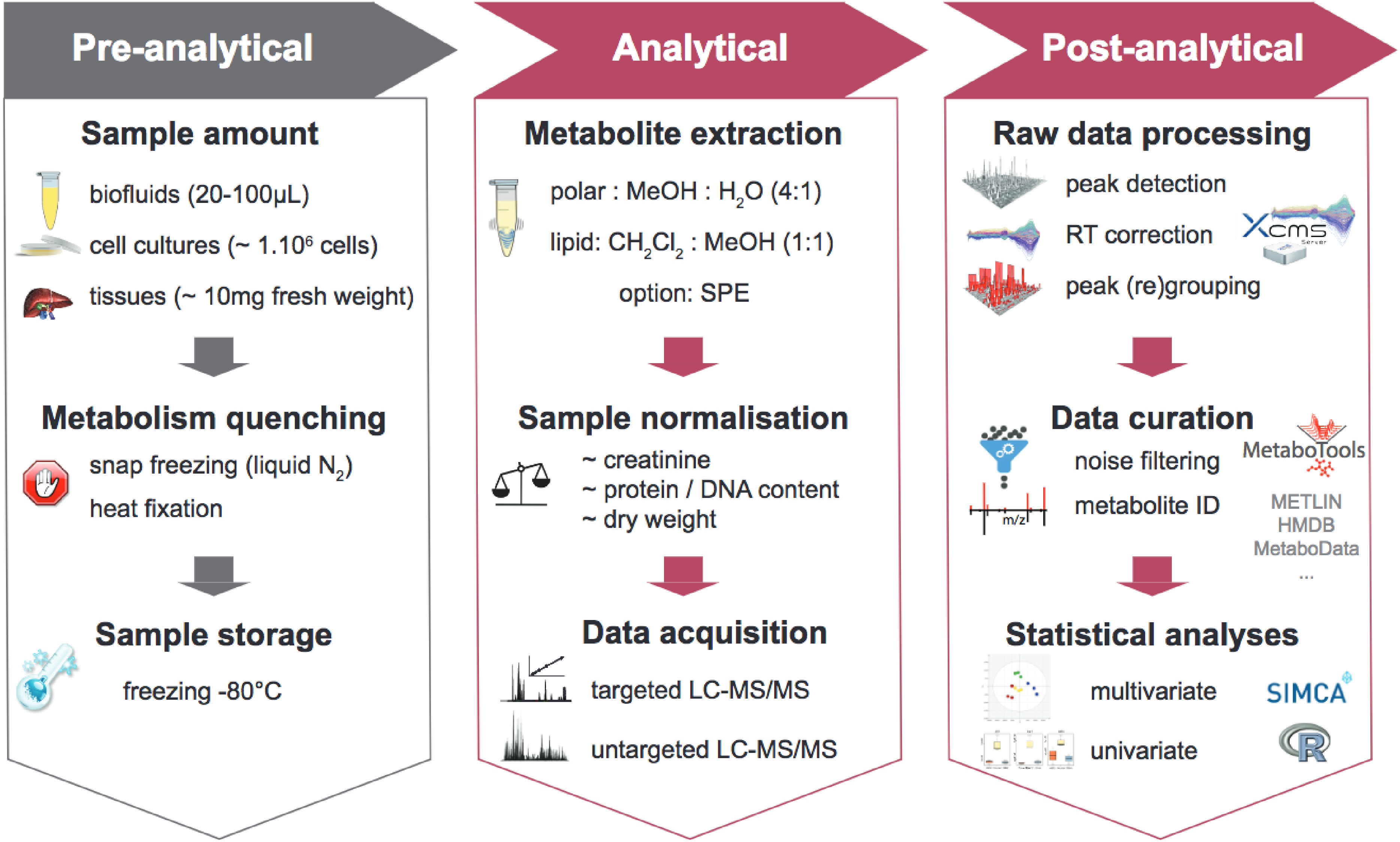
- multivariate
- univariate

SIMCA

R

Experimental Design

Fundamental Metabolomics Advanced



Pre-analytical

Sample amount

biofluids (20-100µL)

cell cultures (~ 1.10^6 cells)

tissues (~ 10mg fresh weight)

Metabolism quenching

snap freezing (liquid N$_2$)

heat fixation

Sample storage

freezing -80°C

Analytical

Metabolite extraction

polar : MeOH : H$_2$O (4:1)

lipid: CH$_2$Cl$_2$: MeOH (1:1)

option: SPE

Sample normalisation

~ creatinine

~ protein / DNA content

~ dry weight

Data acquisition

targeted LC-MS/MS

untargeted LC-MS/MS

Post-analytical

Raw data processing

peak detection

RT correction

peak (re)grouping

Data curation

noise filtering

metabolite ID

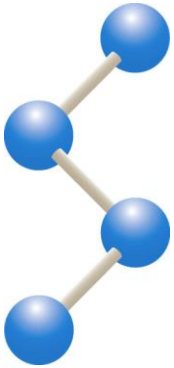
METLIN
HMDB
MetaboData
...

Statistical analyses

multivariate

univariate

Experimental Design Biology



Fundamental Metabolomics

Genomics is a discipline in
which the complete set of **DNA**
within cells or an organism is analyzed.

Fundamental Metabolomics

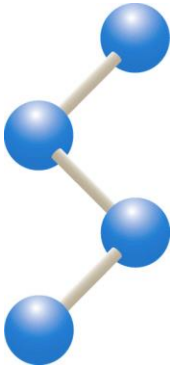
Genomics is a discipline in
which the complete set of **DNA**
within cells or an organism is analyzed.

is a discipline in
which the complete set of
within cells or an organism is analyzed.

Fundamental Metabolomics

Genomics is a discipline in
which the complete set of **DNA**
within cells or an organism is analyzed.

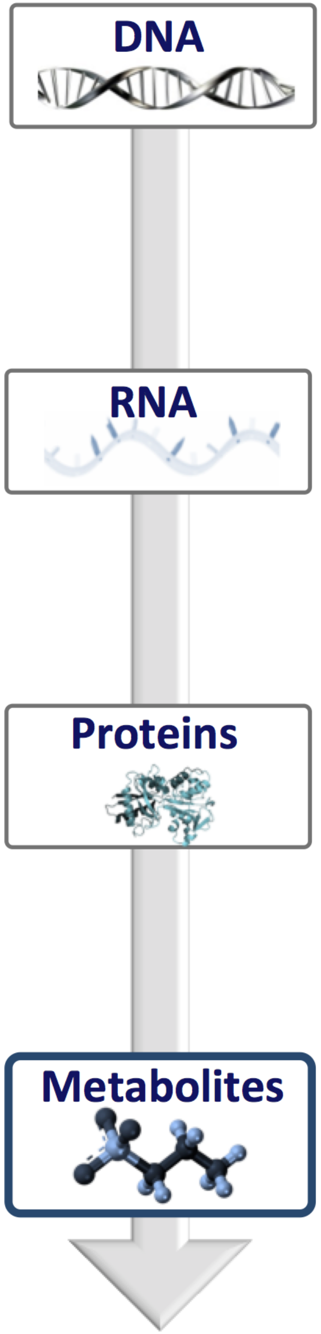
Metabolomics is a discipline in
which the complete set of
within cells or an organism is analyzed.



Fundamental Metabolomics

Genomics is a discipline in
which the complete set of **DNA**
within cells or an organism is analyzed.

Metabolomics is a discipline in
which the complete set of **metabolites**
within cells or an organism is analyzed.



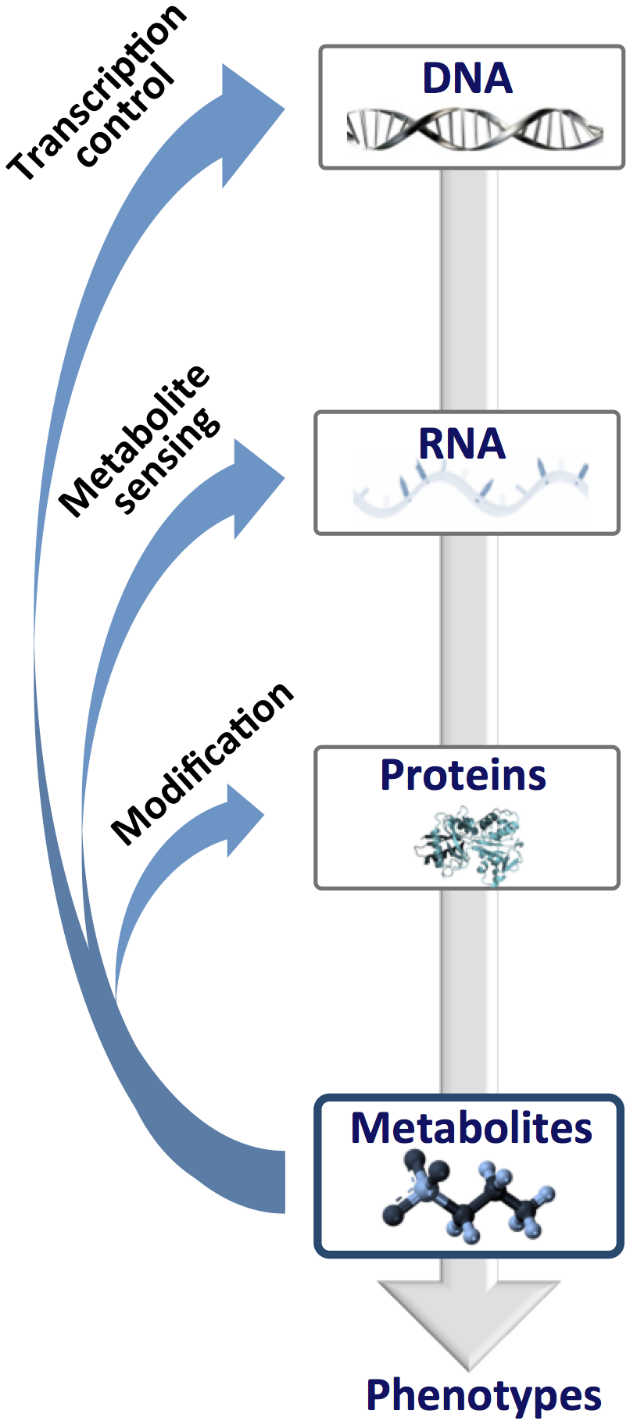
DNA

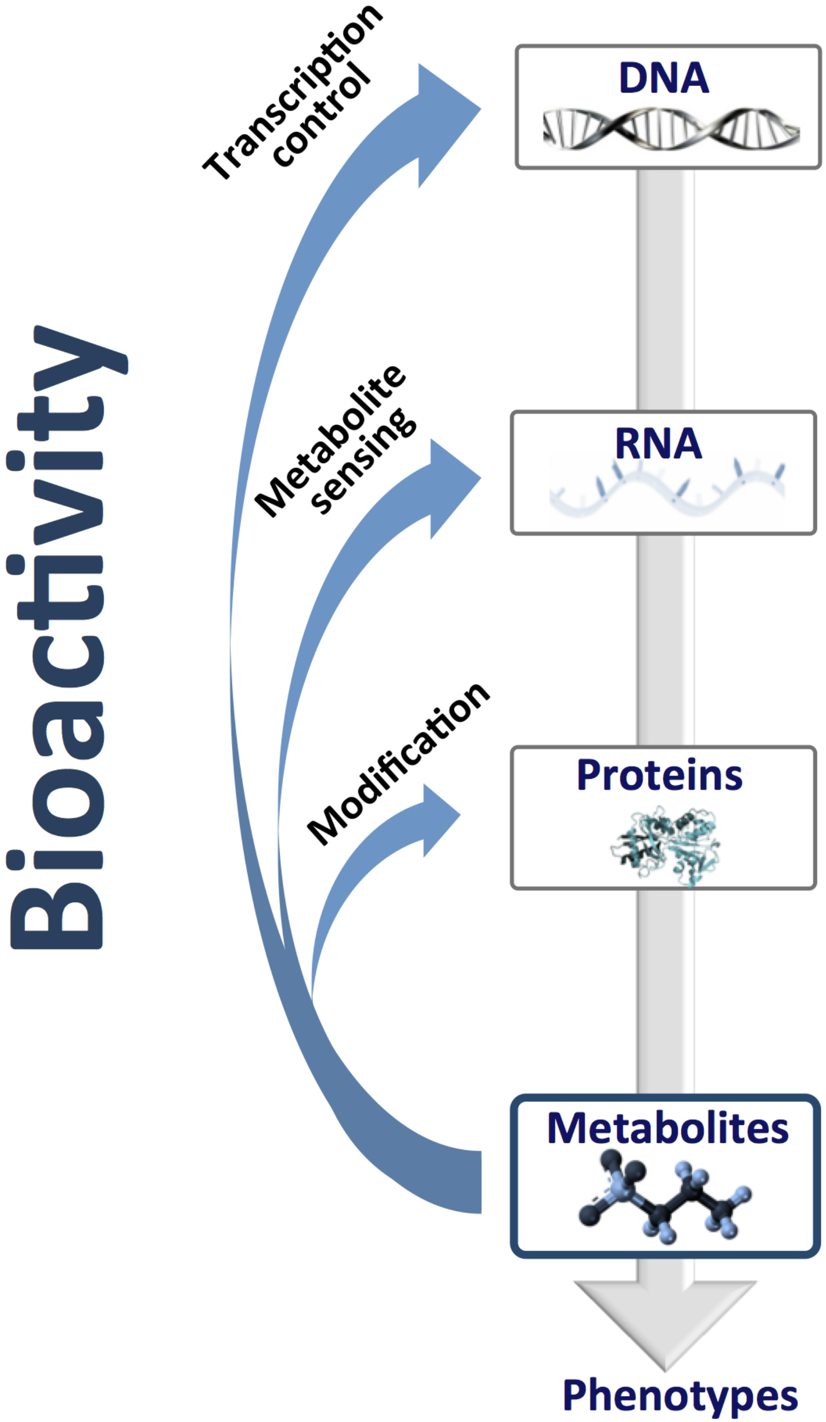
RNA

Proteins

Metabolites

Phenotypes


Transcription control
Metabolite sensing
Modification
DNA
RNA
Proteins
Metabolites
Phenotypes



Bioactivity

Transcription control

Metabolite sensing

Modification

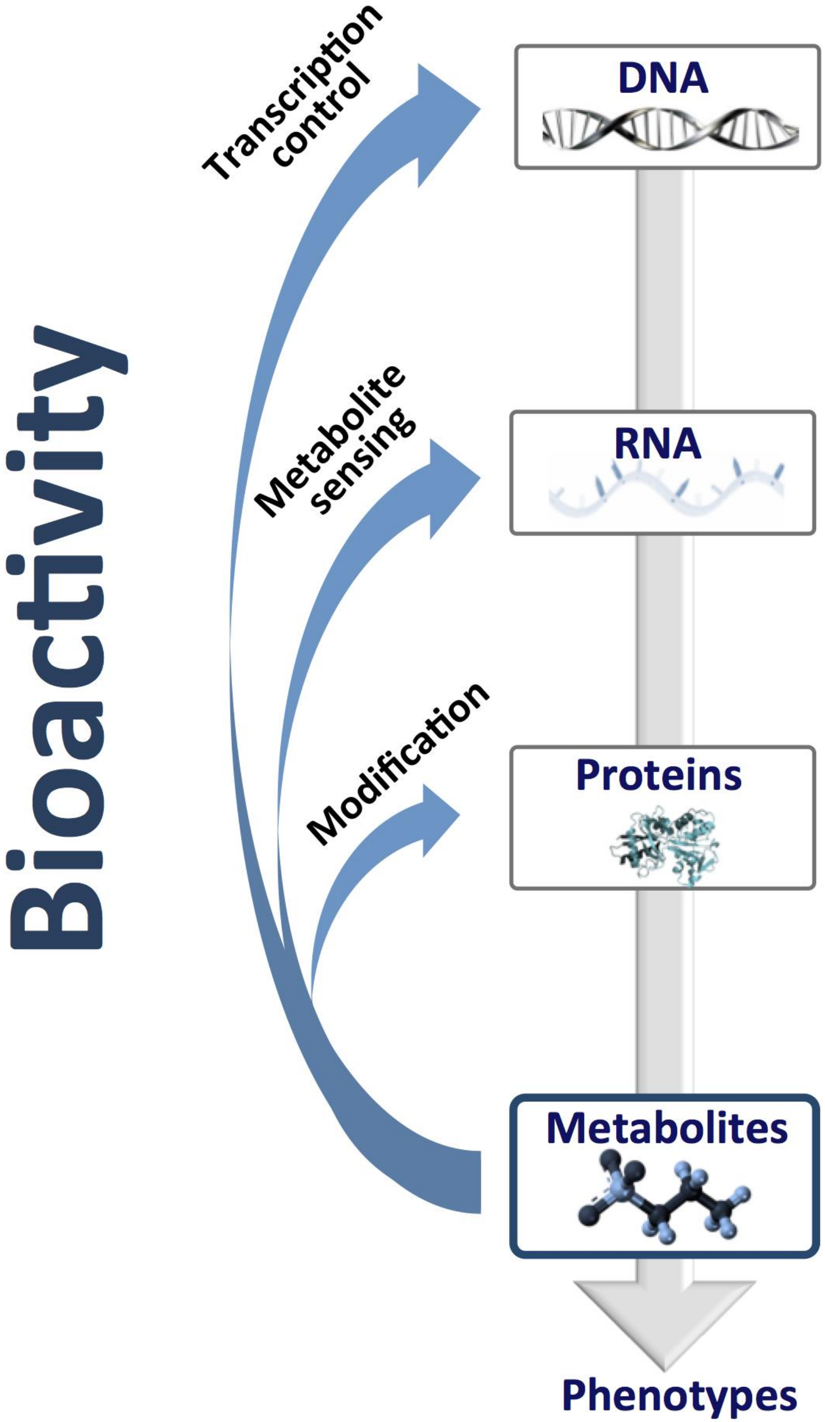
DNA

RNA

Proteins

Metabolites

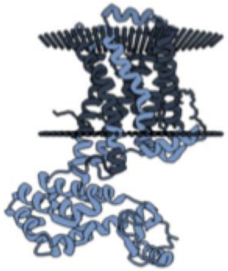
Phenotypes

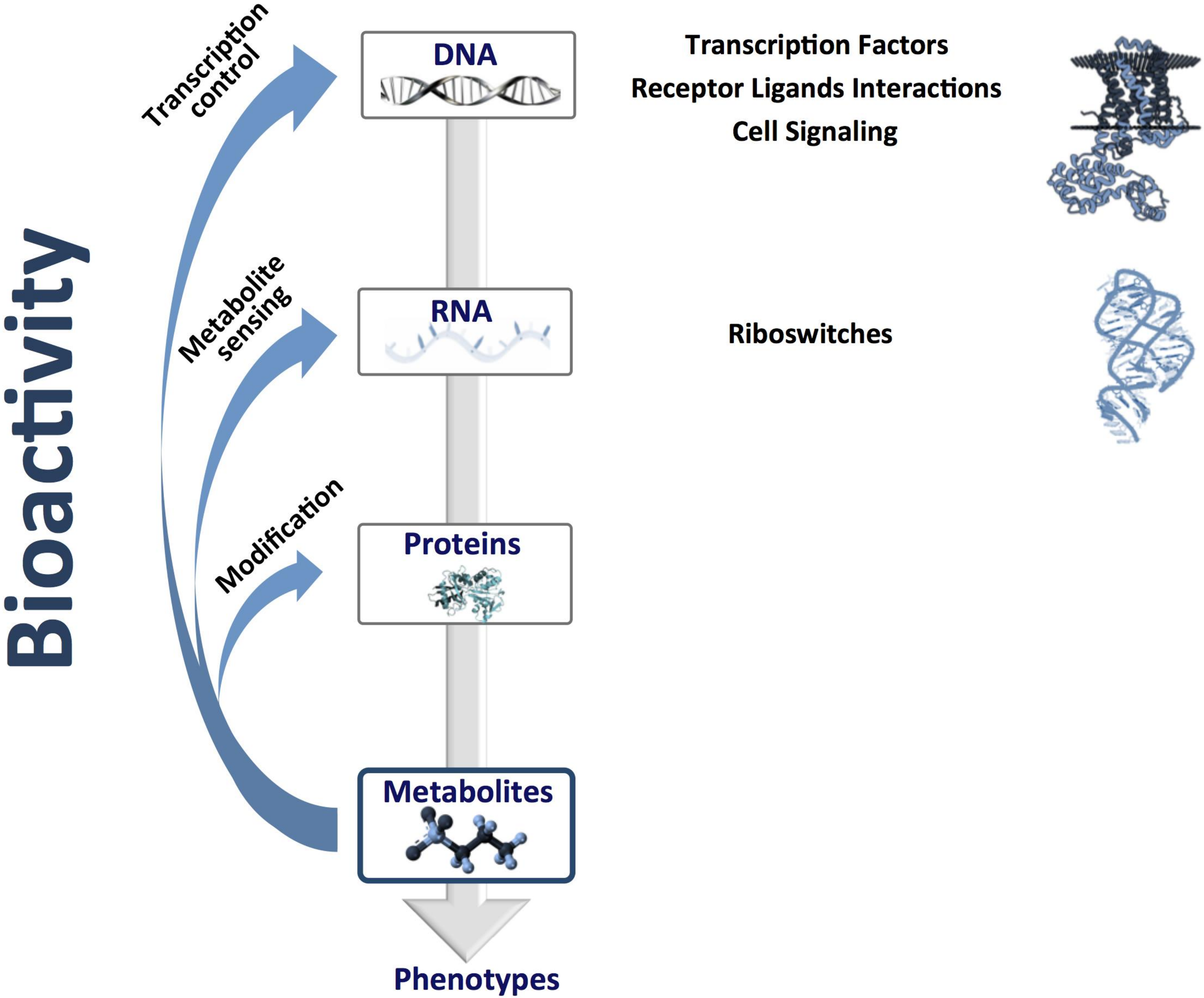


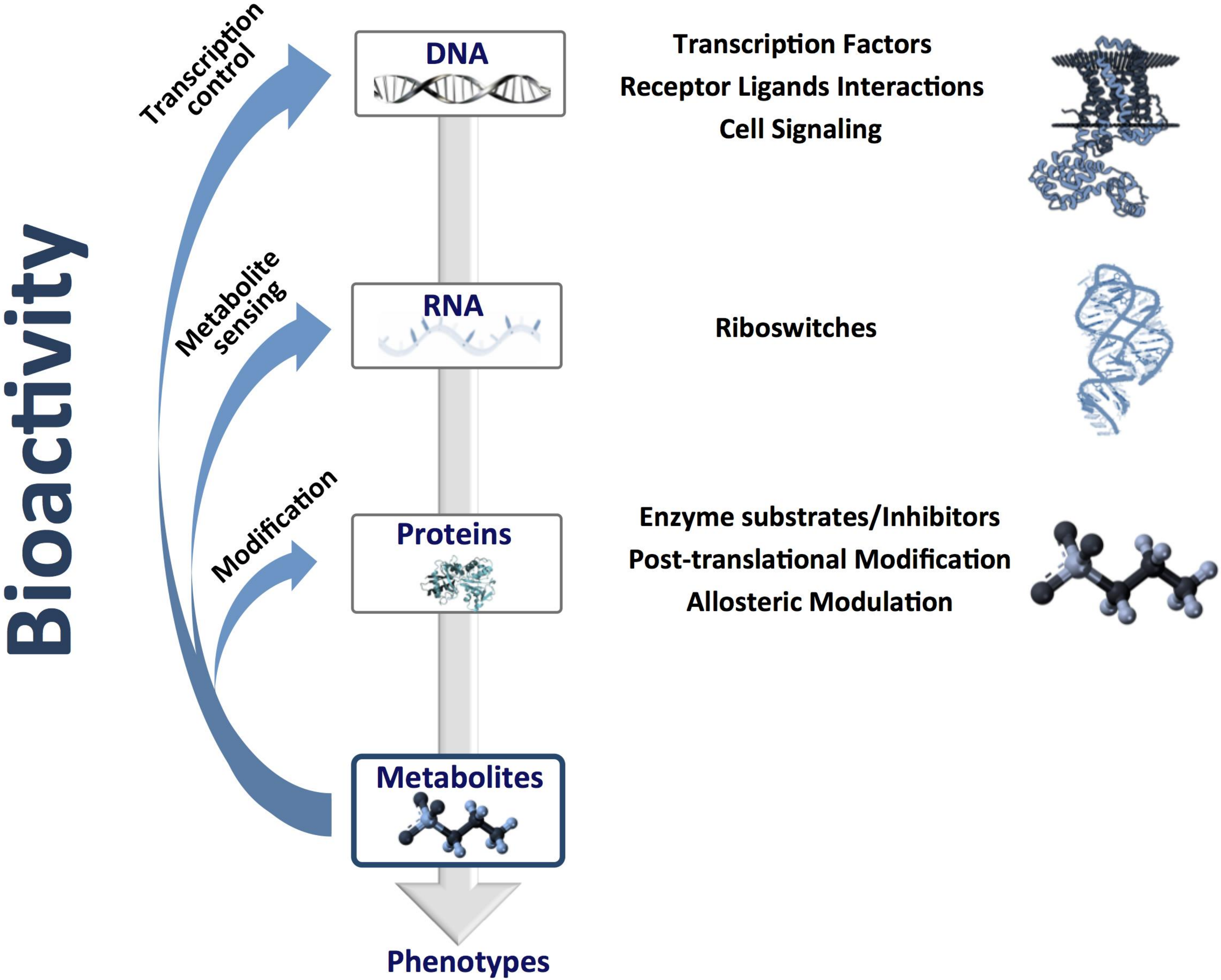
Transcription Factors

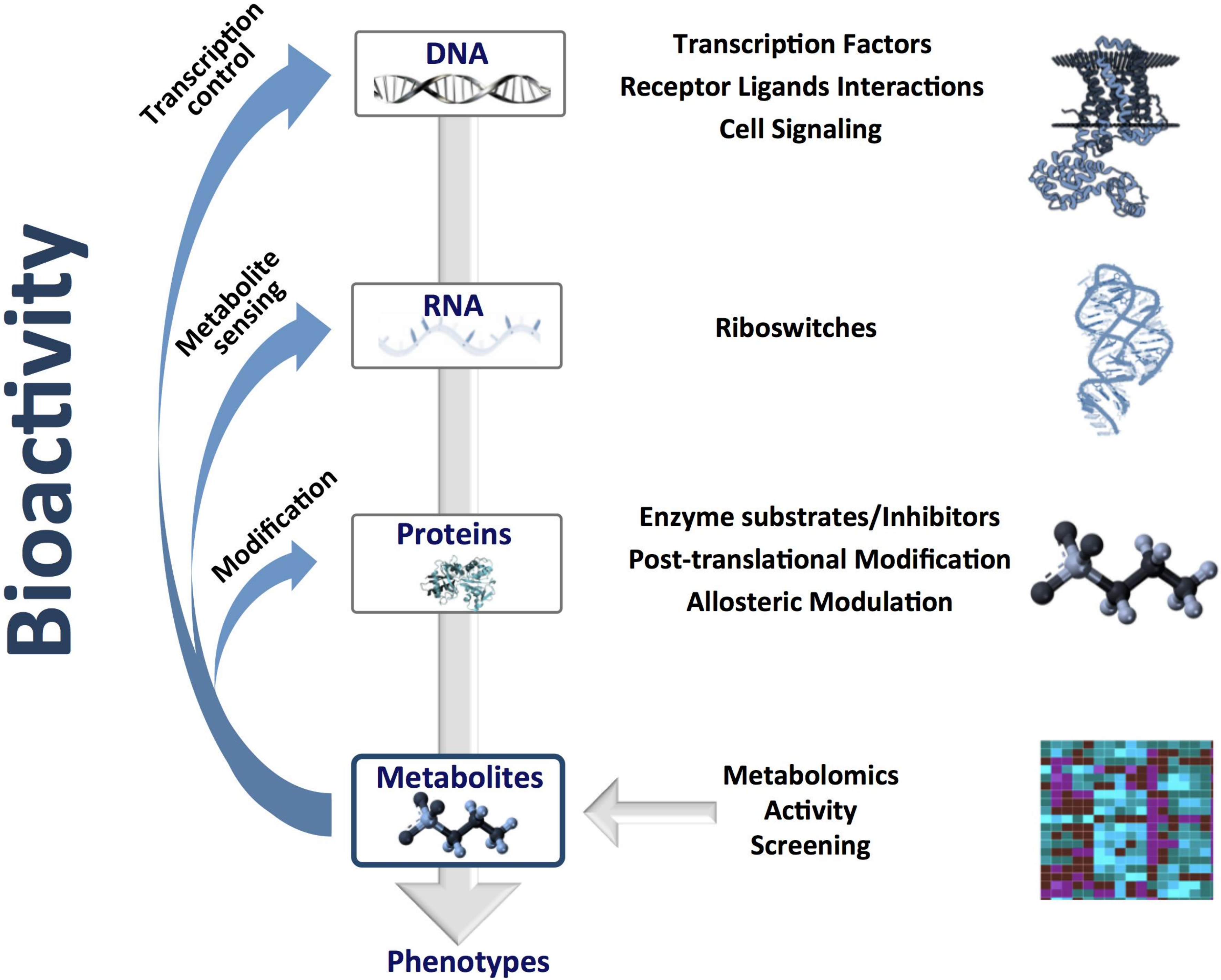
Receptor Ligands Interactions

Cell Signaling

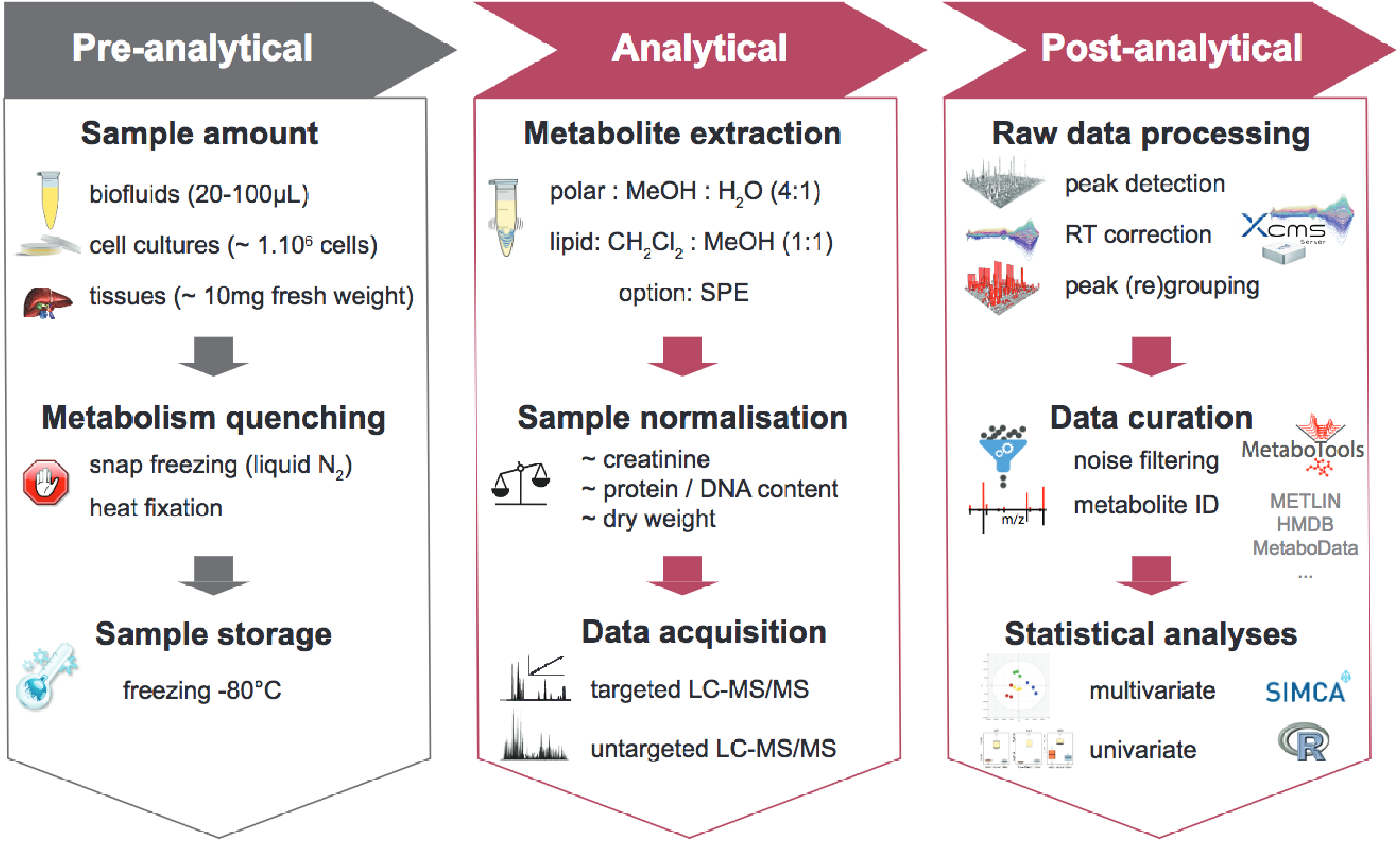



Bioactivity
Transcription control
Metabolite sensing
Modification
DNA
RNA
Proteins
Metabolites
Phenotypes
Transcription Factors
Receptor Ligands Interactions
Cell Signaling
Riboswitches


Bioactivity
Transcription control
Metabolite sensing
Modification
DNA
RNA
Proteins
Metabolites
Phenotypes
Transcription Factors
Receptor Ligands Interactions
Cell Signaling
Riboswitches
Enzyme substrates/Inhibitors
Post-translational Modification
Allosteric Modulation


Bioactivity
Transcription control
Metabolite sensing
Modification
DNA
RNA
Proteins
Metabolites
Phenotypes
Transcription Factors
Receptor Ligands Interactions
Cell Signaling
Riboswitches
Enzyme substrates/Inhibitors
Post-translational Modification
Allosteric Modulation
Metabolomics
Activity
Screening

Fundamental Metabolomics Advanced



Pre-analytical

Sample amount

biofluids (20-100µL)

cell cultures (~ 1.10^6 cells)

tissues (~ 10mg fresh weight)

Metabolism quenching

snap freezing (liquid N_2)

heat fixation

Sample storage

freezing -80°C

Analytical

Metabolite extraction

polar : MeOH : H_2O (4:1)

lipid: CH_2Cl_2 : MeOH (1:1)

option: SPE

Sample normalisation

~ creatinine

~ protein / DNA content

~ dry weight

Data acquisition

targeted LC-MS/MS

untargeted LC-MS/MS

Post-analytical

Raw data processing

peak detection

RT correction

peak (re)grouping

XCMS Server

Data curation

noise filtering

metabolite ID

MetaboTools

METLIN

HMDB

MetaboData

...

Statistical analyses

multivariate

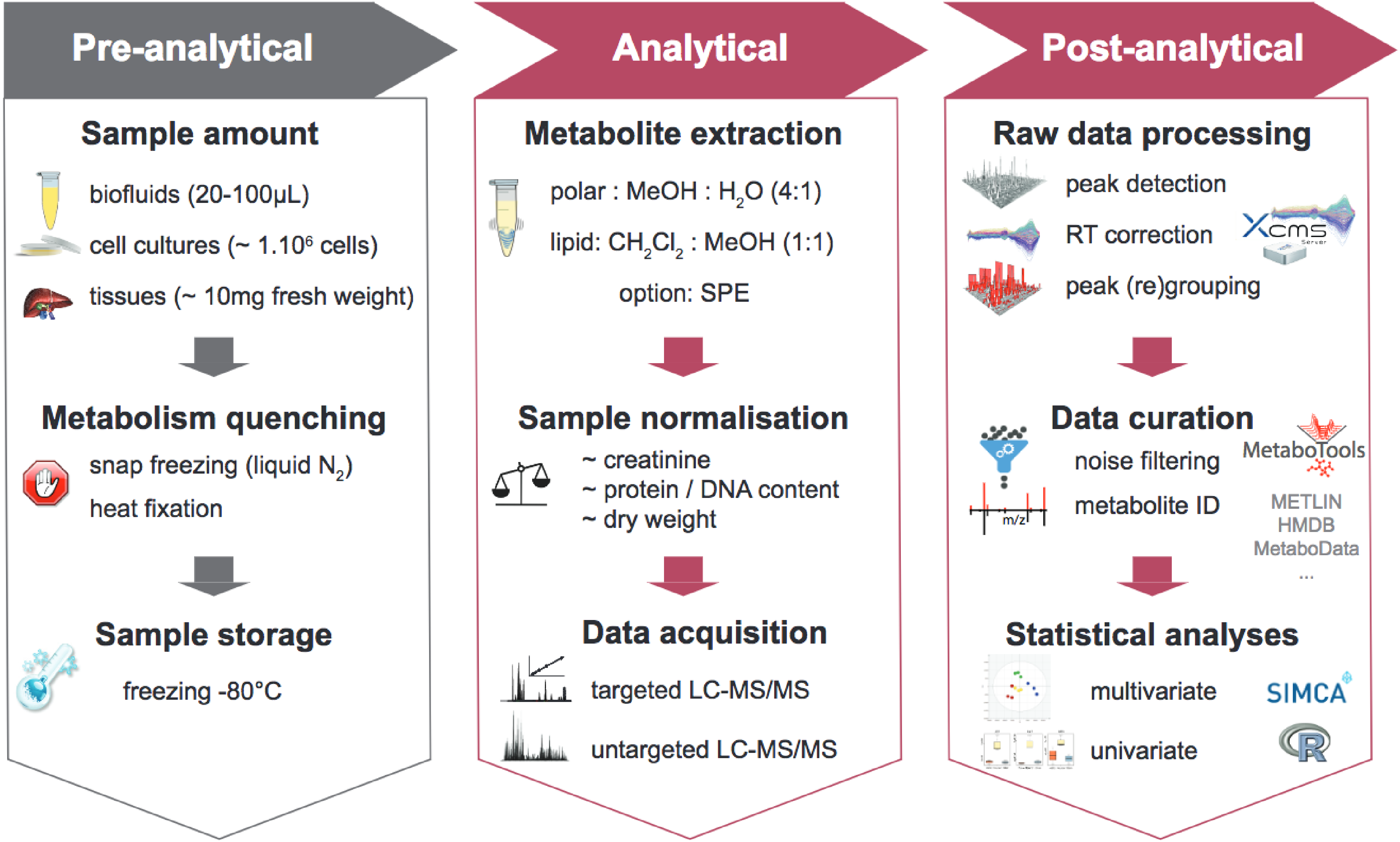
univariate

SIMCA

R

Experimental Design

Fundamental Metabolomics Advanced



Pre-analytical

Sample amount

biofluids (20-100µL)

cell cultures (~ 1.10^6 cells)

tissues (~ 10mg fresh weight)

Metabolism quenching

snap freezing (liquid N_2)

heat fixation

Sample storage

freezing -80°C

Analytical

Metabolite extraction

polar : MeOH : H_2O (4:1)

lipid: CH_2Cl_2 : MeOH (1:1)

option: SPE

Sample normalisation

~ creatinine

~ protein / DNA content

~ dry weight

Data acquisition

targeted LC-MS/MS

untargeted LC-MS/MS

Post-analytical

Raw data processing

peak detection

RT correction

peak (re)grouping

XCMS Server

Data curation

noise filtering

metabolite ID

MetaboTools

METLIN
HMDB
MetaboData
...

Statistical analyses

multivariate

univariate

SIMCA

R

Experimental Design Biology

Fundamental Metabolomics

Untargeted Metabolomics

Experimental Design

Fundamental Metabolomics

Untargeted Metabolomics

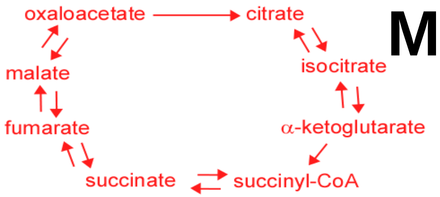
Targeted Metabolomics

Experimental Design

Fundamental Metabolomics

Untargeted Metabolomics

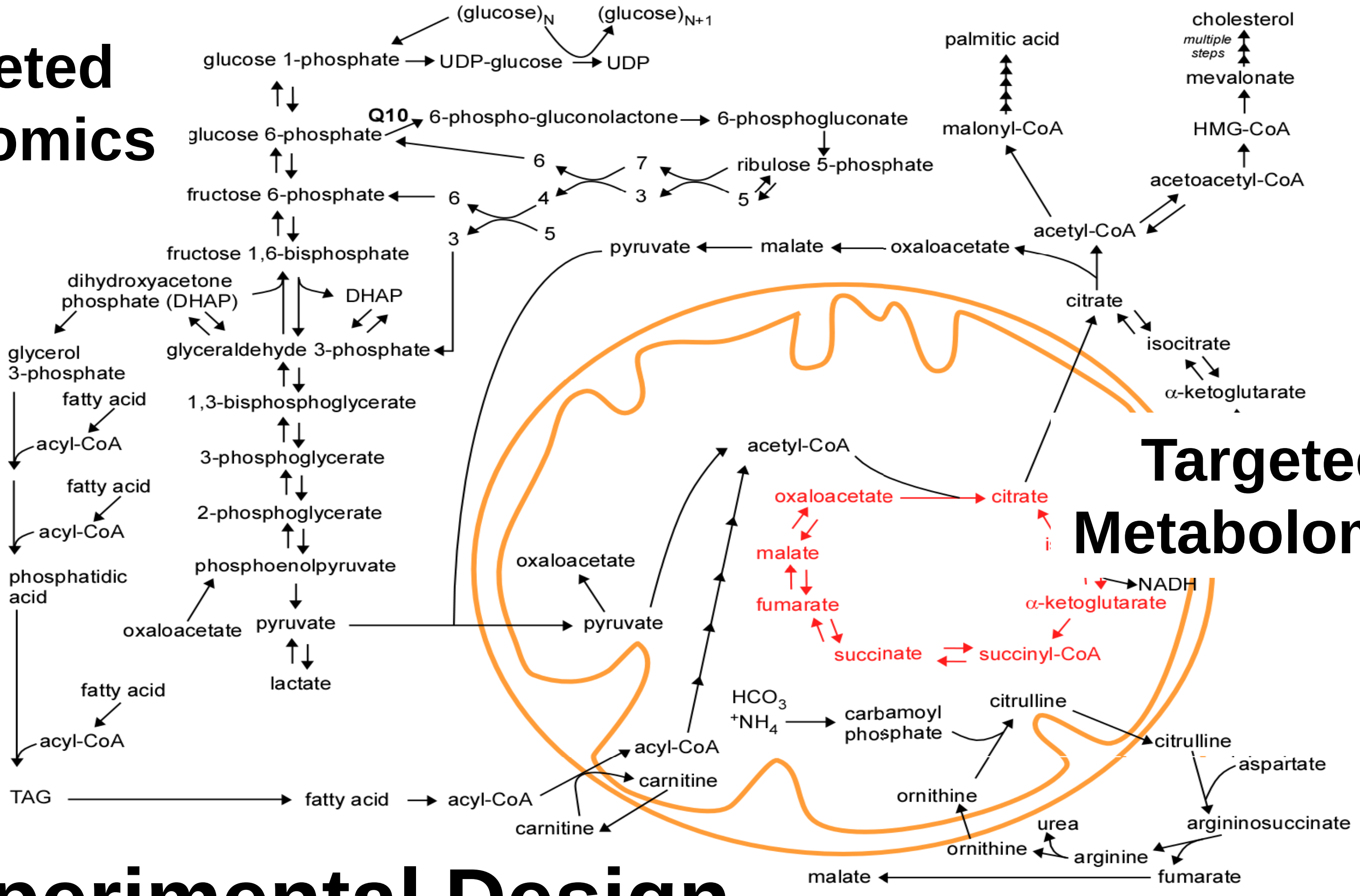
Targeted Metabolomics



Experimental Design

Fundamental Metabolomics

Untargeted Metabolomics



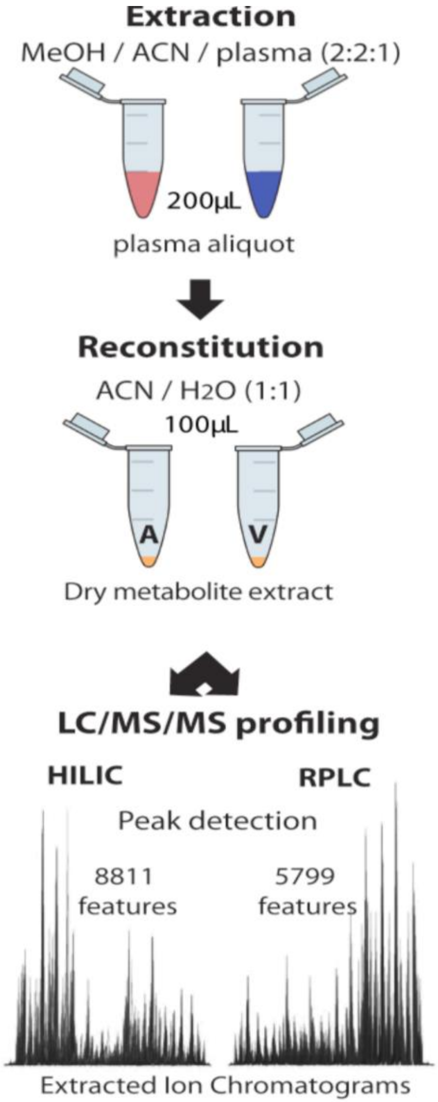
Targeted Metabolomics

Experimental Design

Experimental Design

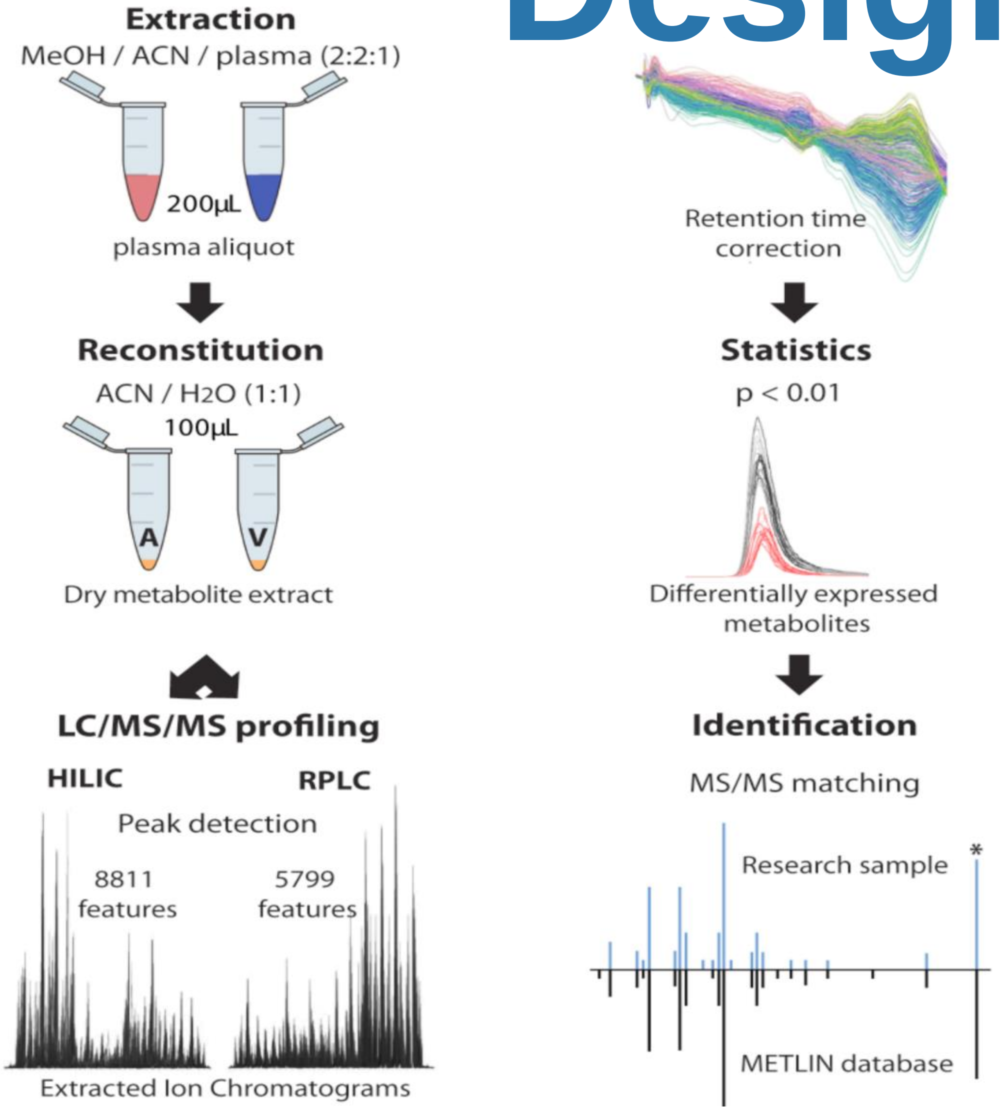
Experimental Design

Untargeted Metabolomics



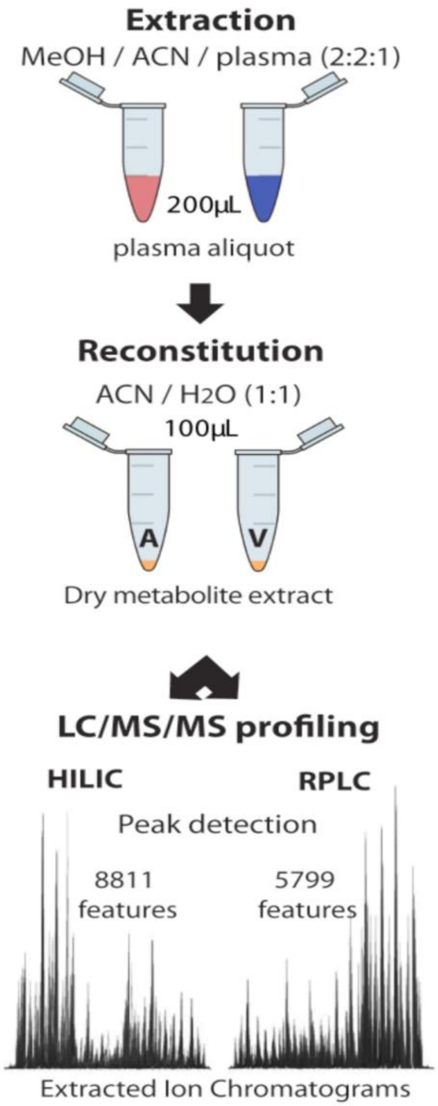
Experimental Design

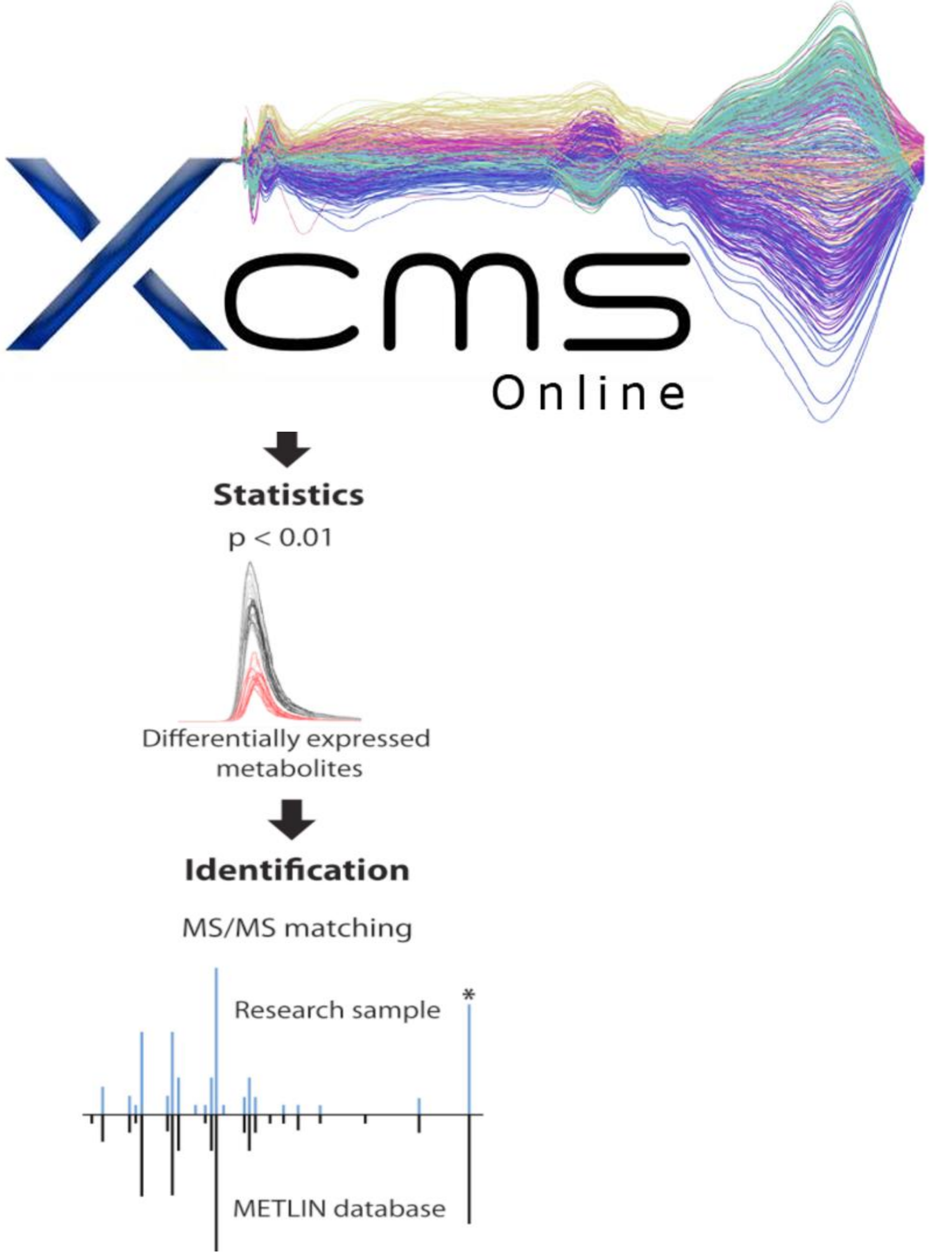
Untargeted Metabolomics



Experimental

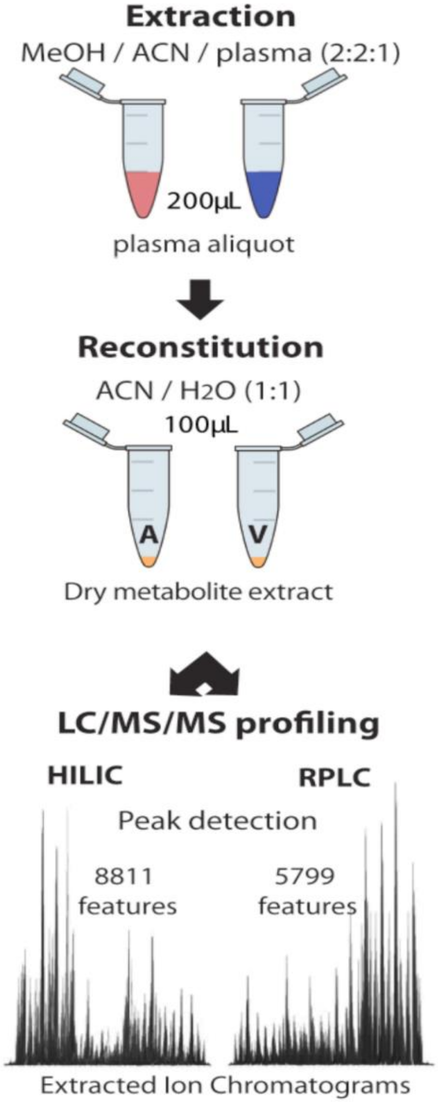
Untargeted Metabolomics





Experimental

Untargeted Metabolomics


Extraction
MeOH / ACN / plasma (2:2:1)

plasma aliquot

Reconstitution
ACN / H_2O (1:1)

Dry metabolite extract

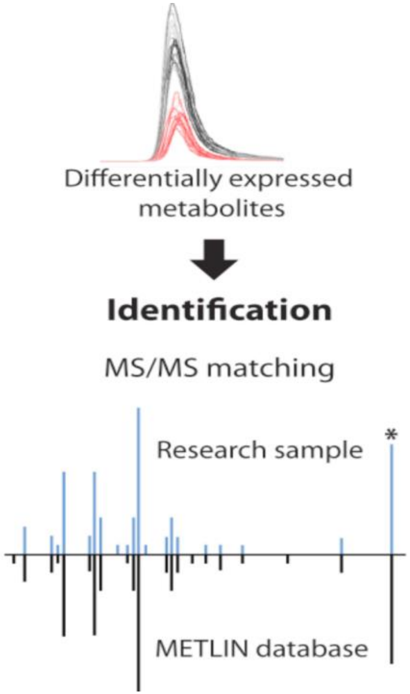
LC/MS/MS profiling

Peak detection

Extracted Ion Chromatograms

xcms Online

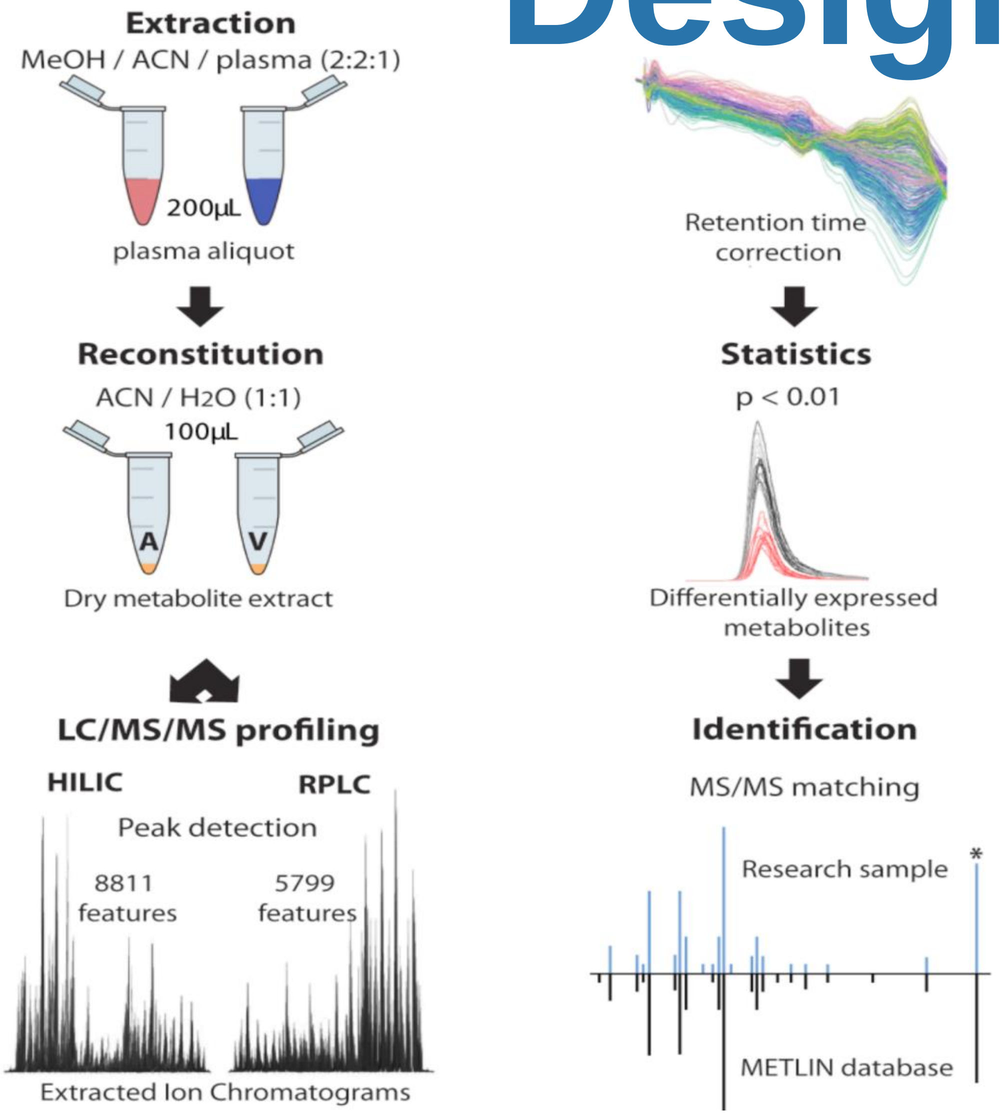
24000 2018-05-16 09:01:14


Differentially expressed metabolites

Identification
MS/MS matching

Experimental Design

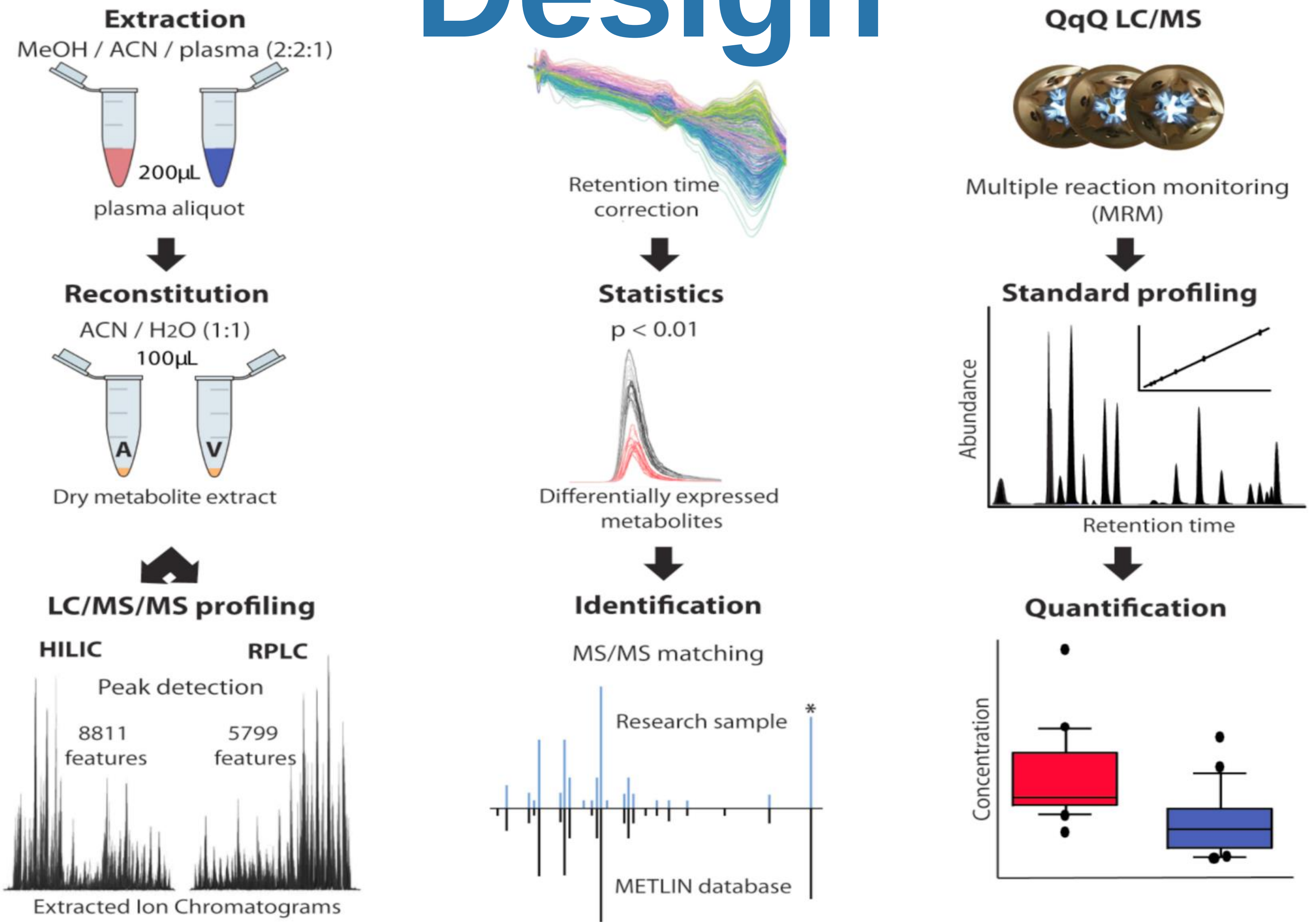
Untargeted Metabolomics



Experimental Design

Untargeted Metabolomics

Targeted Metabolomics



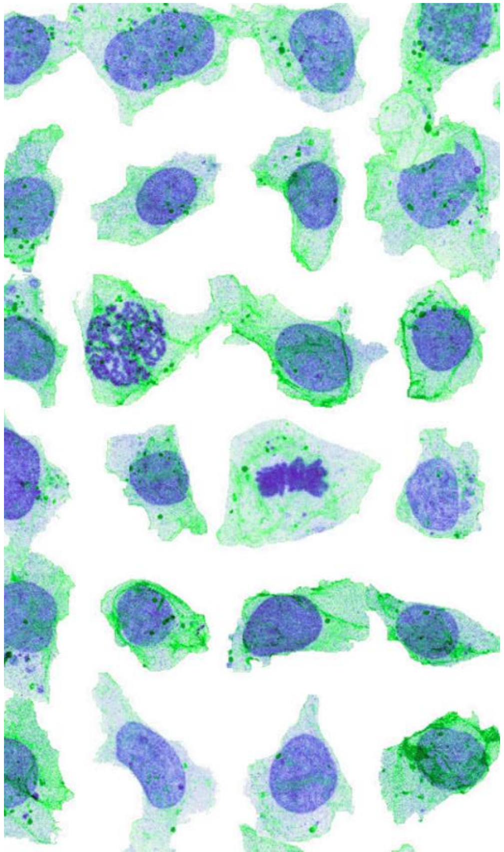
Experimental Design

Experimental Design

Biological Question

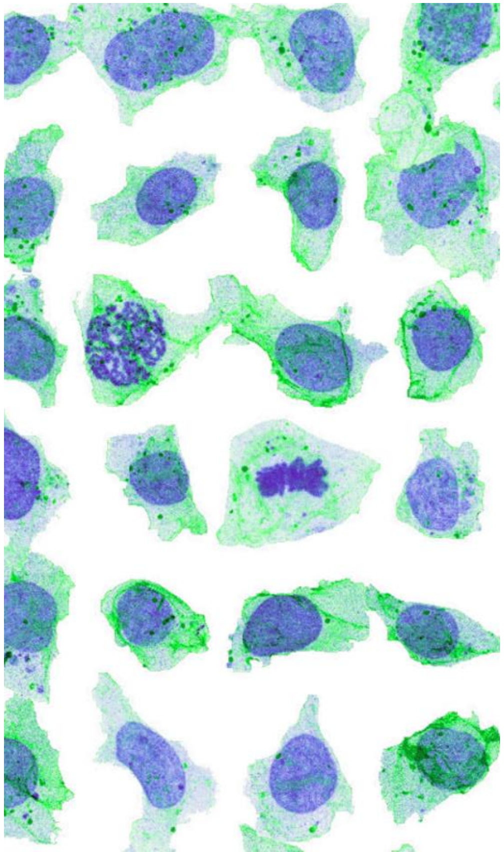
Experimental Design

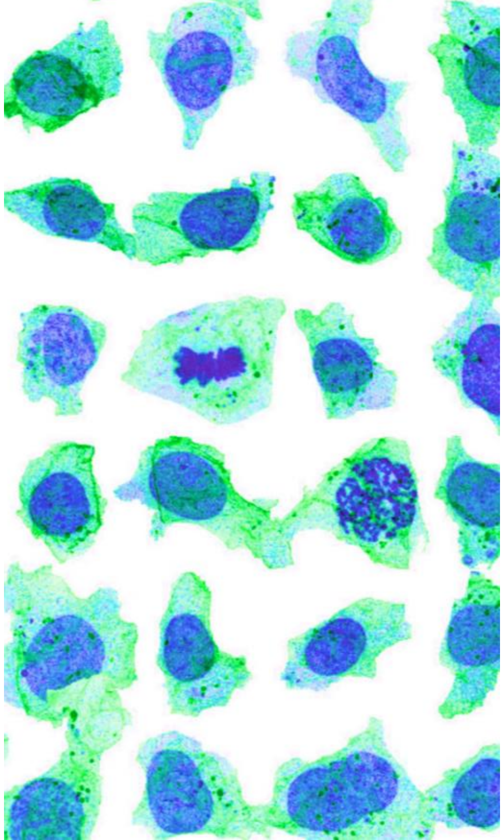
Biological Question



Experimental Design

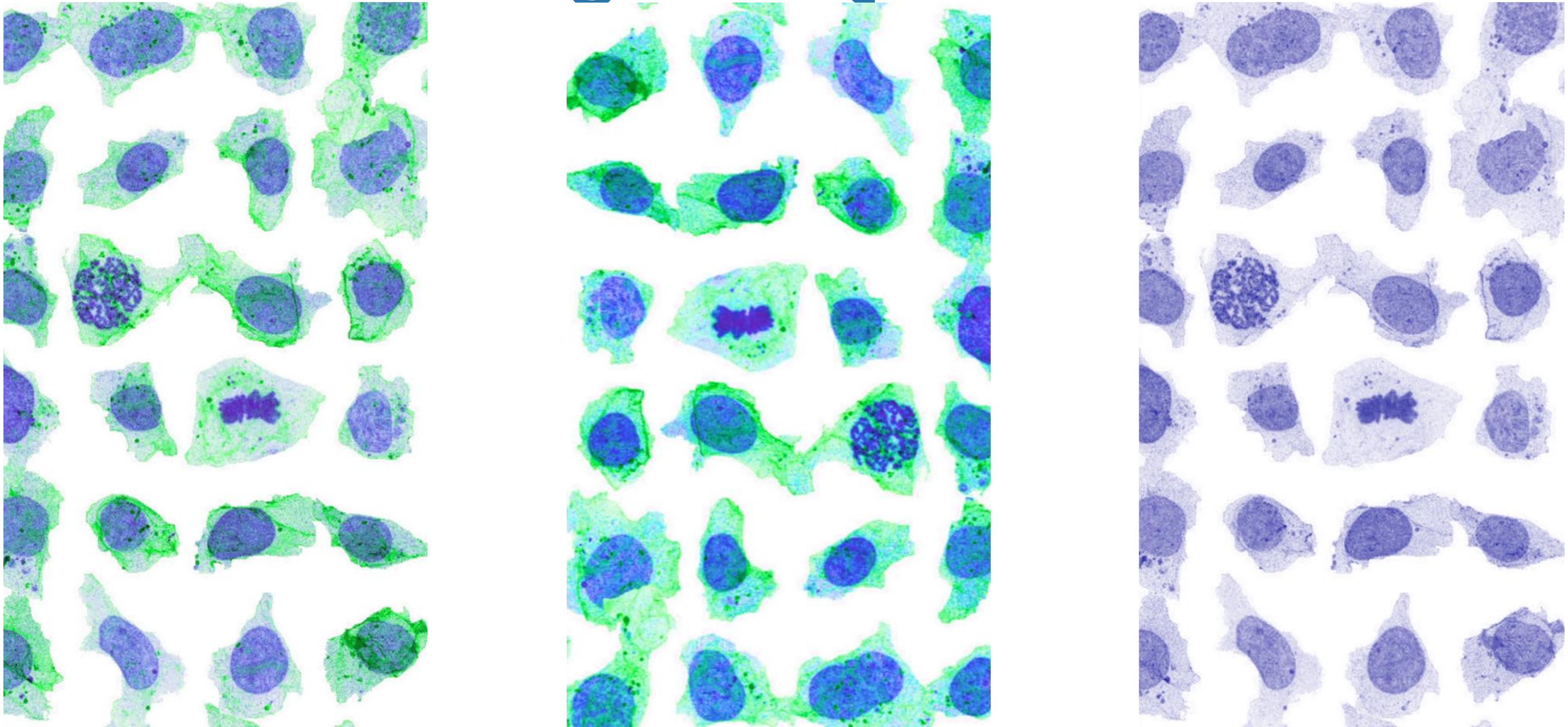
Biological Question





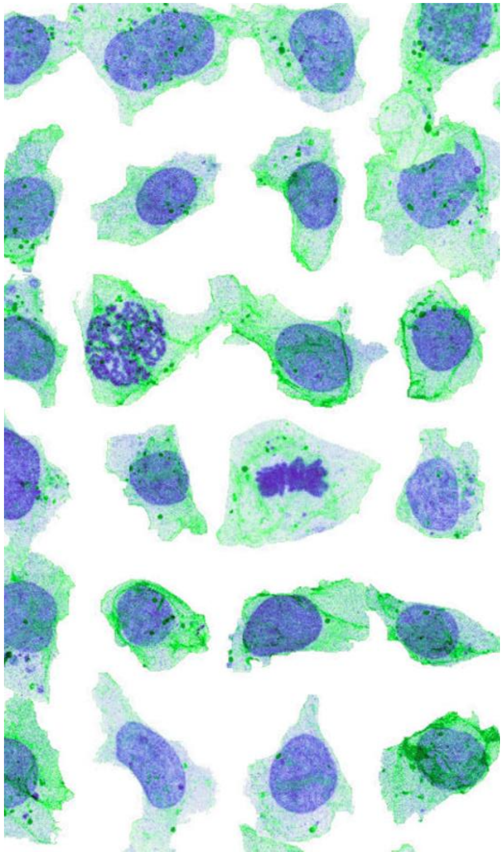
Experimental Design

Biological Question



Experimental Design

Biological Question



Experimental Design

Biological Question

Experimental Design

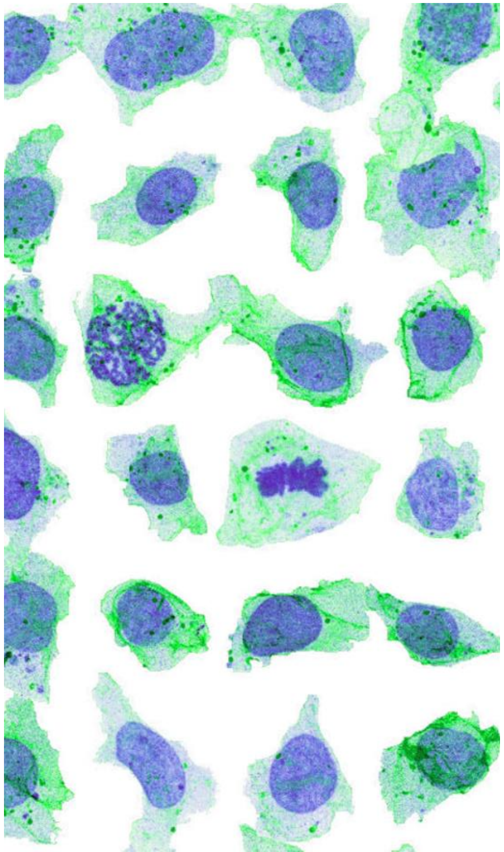
Biological Question

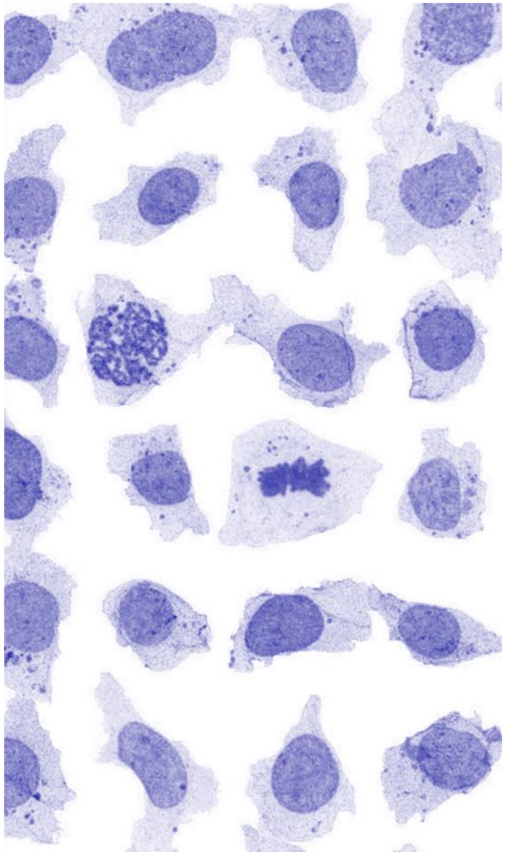
Control Design

Experimental Design

Biological Question

Control Design

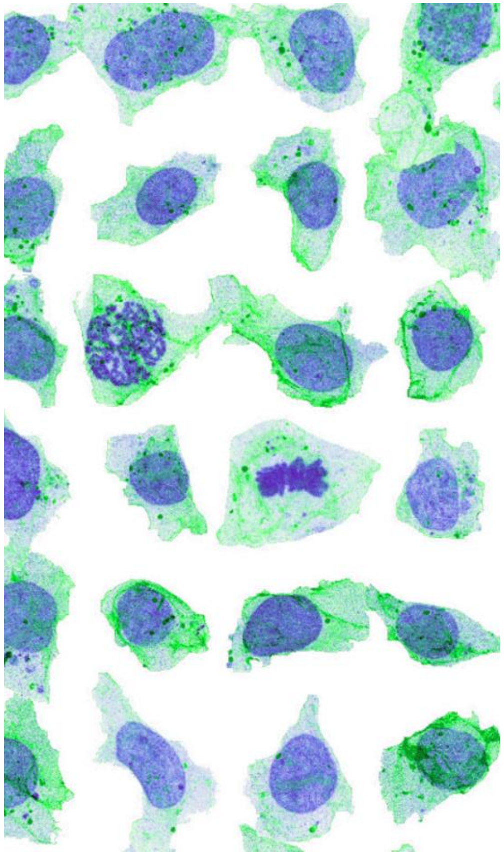


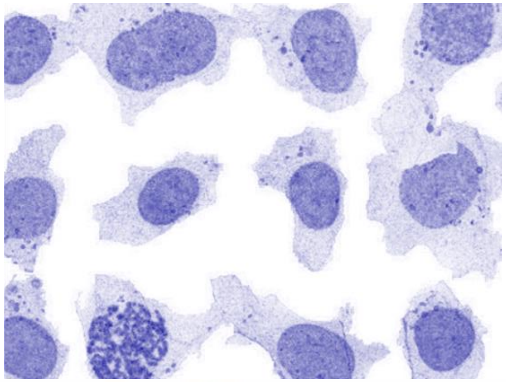


Experimental Design

Biological Question

Control Design





Experimental Design

Biological Question

Control Design

Numbers/Statistics

Experimental Design

Biological Question

Control Design

Numbers/Statistics

Experimental Design

Biological Question

Control Design

Numbers/Statistics

Experimental Design

Biological Question

Control Design

Numbers/Statistics

1 3 5

Experimental Design

Biological Question

Control Design

Numbers/Statistics

1 3 5 10 20

Experimental Design

Biological Question

Control Design

Numbers/Statistics

1 3 5 10 20 50

Experimental Design

Biological Question

Control Design

Numbers/Statistics

1 3 5 10 20 50 100 1000

Experimental Design

Biological Question

Control Design

Numbers/Statistics

Biomarkers/Pathways/Mechanism

Experimental Design

Biological Question

Control Design

Numbers/Statistics

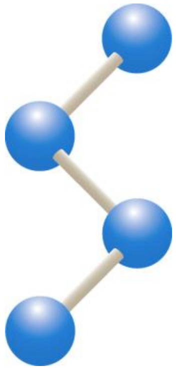
Biomarkers/Pathways/Mechanism

Biological Activity

Metabolomics Papers PDFs Included

Metabolomics Activity Screening Identifies Metabolites that Modulate Phenotype	**Nature Biotechnology 2018**
Data Processing, Multi-Omic Pathway Mapping, and Metabolite Activity Analysis	**Nature Protocols 2018**
Metabolomics-Based Discovery of a Metabolite that Enhances Oligodendrocyte Maturation	**Nature Chemical Biology 2018**
Metabolite-Induced Protein Expression Guided by Metabolomics and Systems Biology	**Cell Metabolism 2018**
Systems Biology guided by Metabolomics	**Nature Methods 2017**
Metabolomics: Beyond Biomarkers and Towards Mechanisms	**Nature Reviews 2016**
XCMS: Processing MS Data using Nonlinear Alignment and Metabolite ID	**Analytical Chemistry 2006**
Mzmine 2: Modular framework for processing, visualizing, and analyzing MS data	**BMC Bioinformatics 2010**
Bioinformatics: The Next Frontier of Metabolomics	**Analytical Chemistry 2015**
Predicting Network Activity from High Throughput Metabolomics	**PLOS Computational Biol. 2013**
Interactive XCMS Online: Simplifying Advanced Data Processing and Statistical	**Analytical Chemistry 2014**
Autonomous Metabolomics for Rapid Metabolite Identification in Global Profiling	**Analytical Chemistry 2015**
Thermal Degradation of Small Molecules: A Global Metabolomics Investigation	**Analytical Chemistry 2015**
Arteriovenous Blood Metabolomics: A Readout of Intra-Tissue Metabostasis	**Scientific Reports 2015**
Metabolism Links Bacterial Biofilms and Colon Carcinogenesis	**Cell Metabolism 2015**
CFM-ID: a web server for annotation, prediction and metabolite ID	N**ucleic Acid Research 2014**
Determining Conserved Metabolic Biomarkers from a Million Database Queries	**Bioinformatics 2015**
Autonomous Metabolomics for Rapid Metabolite Identification in Global Profiling	**Analytical Chemistry 2015**
Metabolomic data streaming for biology-dependent data acquisition	**Nature Biotechnology 2014**
Comprehensive bioimaging with fluorinated nanoparticles	**Nature Comm. 2015**
Liquid chromatography quadrupole time-of-flight mass spectrometry	**Nature Protocols 2013**
Multivariate Analysis in Metabolomics	**Current Metabolomics 2013**
Intra- and Interlaboratory Reproducibility of UPLC TOF MS for Urinary Metabolic Profiling	**Analytical Chemistry 2012**
A Guideline to Univariate Statistical Analysis for LC/MS	**Metabolites 2012**
An accelerated workflow for untargeted metabolomics using METLIN database	**Nature Biotechnology 2012**
Within-Day Reproducibility of an HPLC-MS-Based Method	**Journal Proteome Research 2007**
HMDB: the Human Metabolome Database	**Nucleic Acid Research 2007**
XCMS: Processing MS Data using Nonlinear Alignment and Metabolite ID	**Analytical Chemistry 2006**
METLIN: A Mass Spectral Database	**Therap. Drug Monitoring 2005**

May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Statistical Analysis***

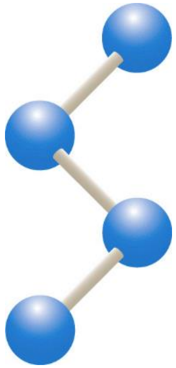
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

May 16th



Fundamental Metabolomics

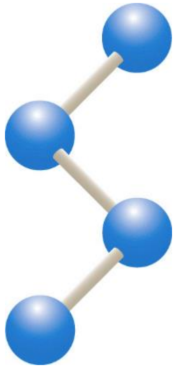
- *Objectives and Challenges*
- *Experimental Design*
- *Sample Preparation and Chromatography*
- *Untargeted Metabolomics*
- *Targeted Metabolomics*
- *Statistical Analysis*

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----



Fundamental Metabolomics

May 16th

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Statistical Analysis***

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

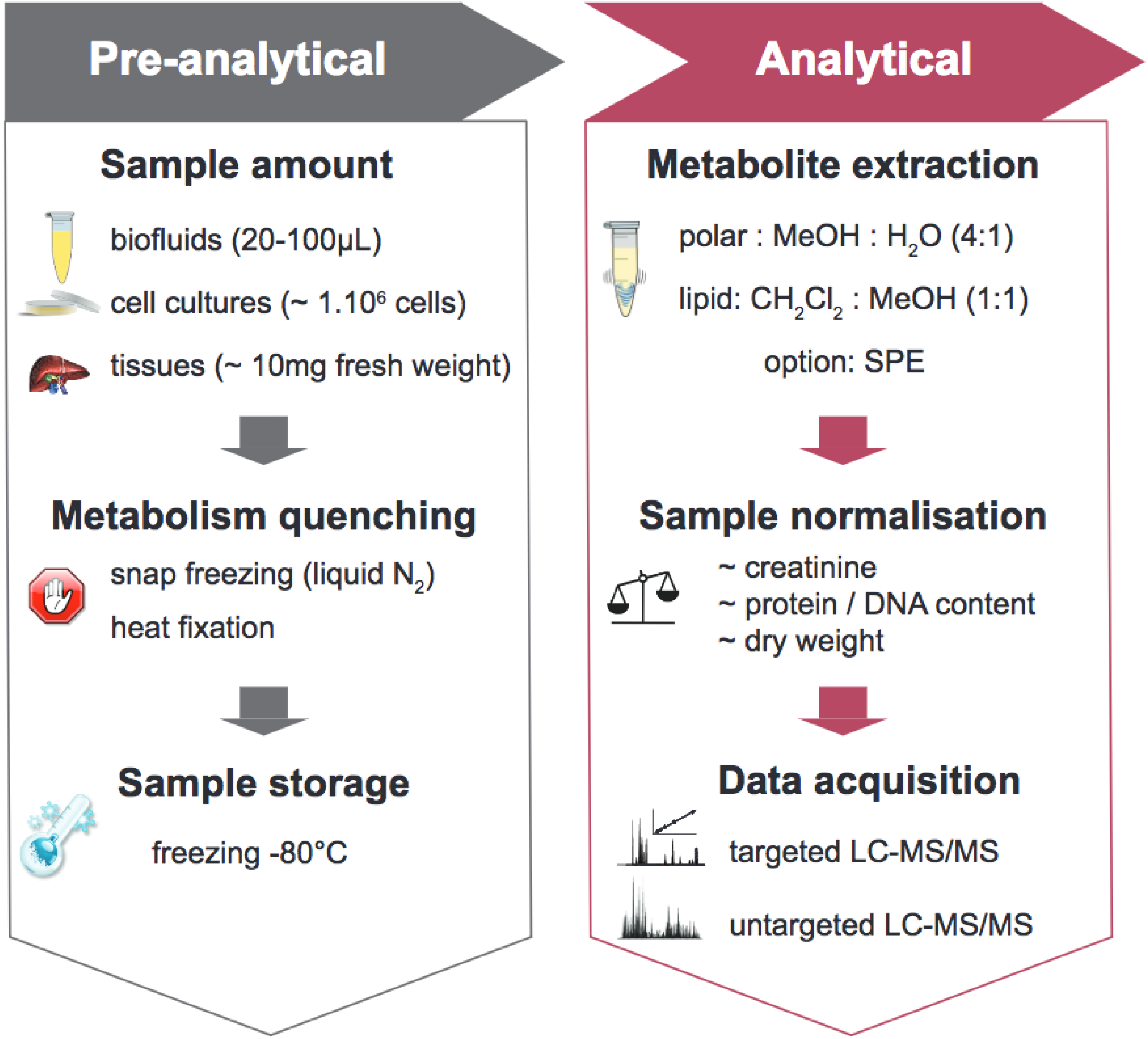
---- 04:30 pm Finish ----

Sample Preparation

May 16th

Elizabeth Billings

Fundamental Metabolomics



Pre-analytical

Sample amount

biofluids (20-100µL)

cell cultures (~ 1.10^6 cells)

tissues (~ 10mg fresh weight)

Metabolism quenching

snap freezing (liquid N_2)

heat fixation

Sample storage

freezing -80°C

Analytical

Metabolite extraction

polar : MeOH : H_2O (4:1)

lipid: CH_2Cl_2 : MeOH (1:1)

option: SPE

Sample normalisation

~ creatinine

~ protein / DNA content

~ dry weight

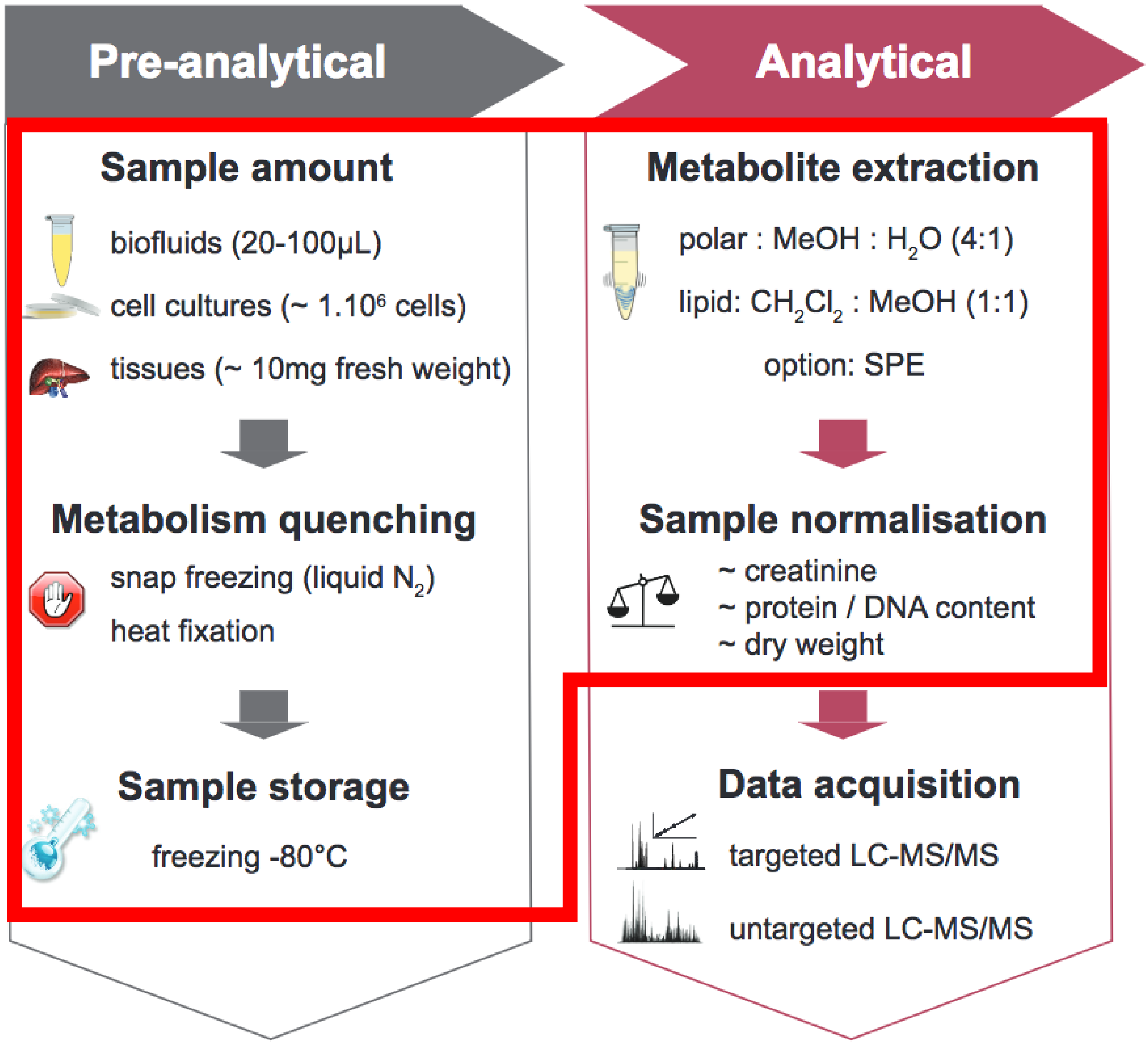
Data acquisition

targeted LC-MS/MS

untargeted LC-MS/MS

Experimental Design

Fundamental Metabolomics



Pre-analytical

Sample amount

biofluids (20-100µL)

cell cultures (~ 1.10^6 cells)

tissues (~ 10mg fresh weight)

Metabolism quenching

snap freezing (liquid N_2)

heat fixation

Sample storage

freezing -80°C

Analytical

Metabolite extraction

polar : MeOH : H_2O (4:1)

lipid: CH_2Cl_2 : MeOH (1:1)

option: SPE

Sample normalisation

~ creatinine

~ protein / DNA content

~ dry weight

Data acquisition

targeted LC-MS/MS

untargeted LC-MS/MS

Experimental Design

Sample Types

Biofluids (20-100µL)

- urine, cerebrospinal fluid (CSF), blood plasma/serum, saliva, etc.

Cell Cultures (varies by cell type, typically 1x10^6 cells)

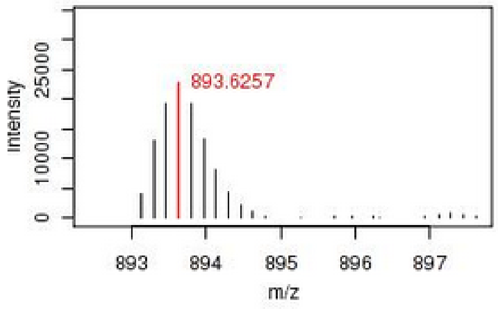
- adherent and non-adherent cells

Tissues (~10 mg fresh weight)

- liver, kidney, muscle, fat, etc.

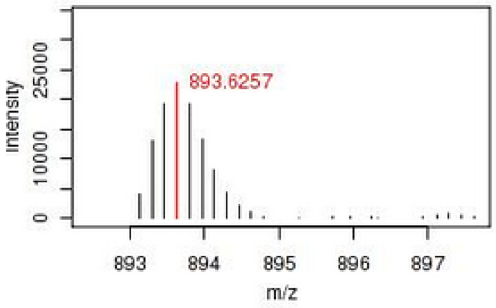
Metabolite Extraction

- Remove proteins and particulates



Metabolite Extraction

- Remove proteins and particulates



- Quench metabolism

Metabolite Extraction

4 approaches:

- Solvent
- Temperature
- pH
- Molecular weight cut-off

Metabolite Extraction

4 approaches:

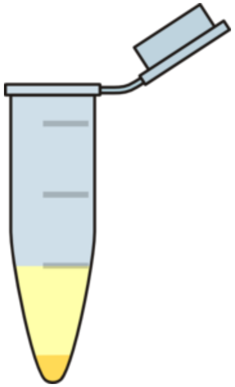
Solvent

Temperature

pH

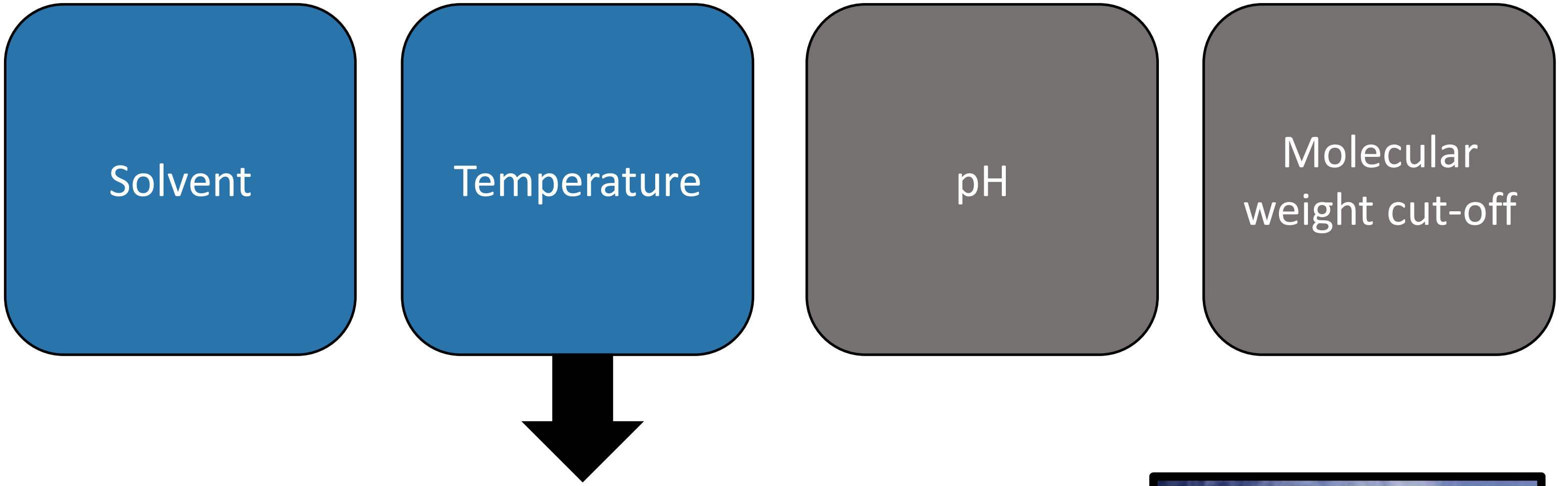
Molecular weight cut-off

- Rapid drop of temperature in solvent and efficient protein precipitation
- Extraction of polar and non-polar molecules
- Polar solvent
  - MeOH, EtOH, H_2O, MeOH:H_2O (4:1)
- Non-polar solvent
  - Acetone, chloroform, hexane



Metabolite Extraction

4 approaches:



Solvent

Temperature

pH

Molecular weight cut-off

- Cold is recommended
  - liquid nitrogen or dry ice ethanol bath
  - very fast quenching
  - much safer for metabolites and metabolomics than heat fixation

Metabolite Extraction

4 approaches:

Solvent

Temperature

pH

Molecular weight cut-off

- Fast inactivation of metabolism, efficient protein precipitation
- Not all metabolites are stable on pH extremes!
  - Acid-stable compounds (nucleotides, water-soluble)
    - Perchloric acid
  - Alkali-stable compounds
    - Ammonium acetate

Metabolite Extraction

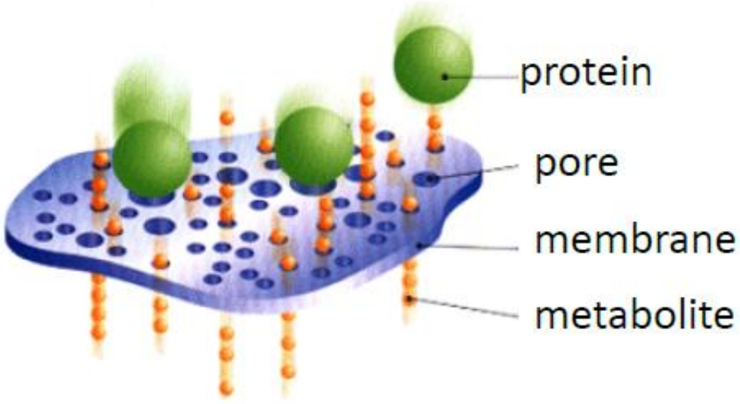
4 approaches:

Solvent

Temperature

pH

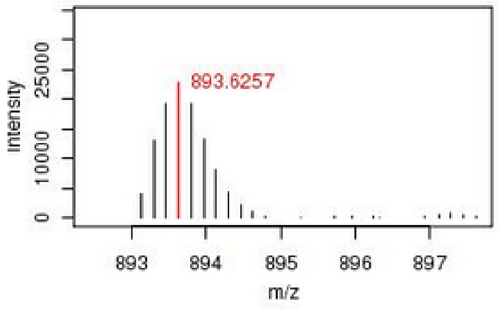
Molecular weight cut-off



- For removal of biomacromolecules
- Metabolites 80 > Da < 1500
- Proteins, DNA, RNA

Metabolite Extraction

- Remove proteins and particulates

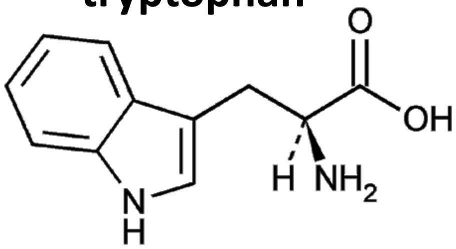


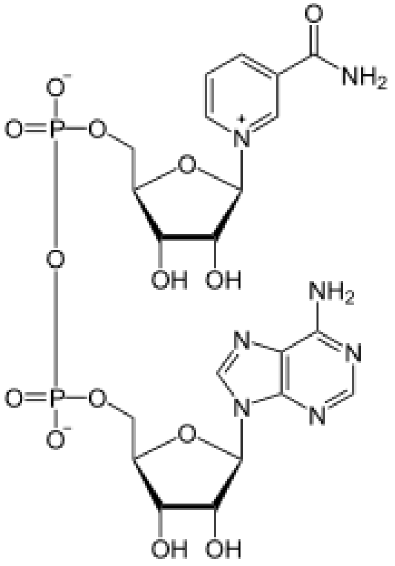
- Quench metabolism

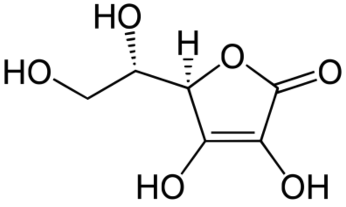
- Recover hydrophobic and hydrophilic metabolites

Metabolite Extraction

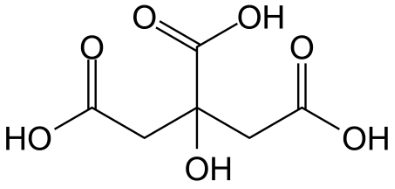
tryptophan



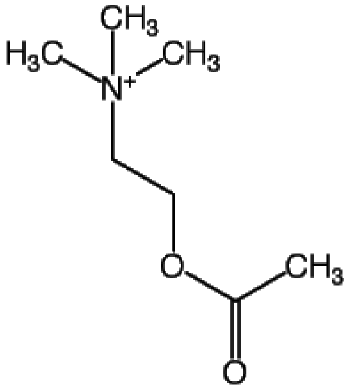


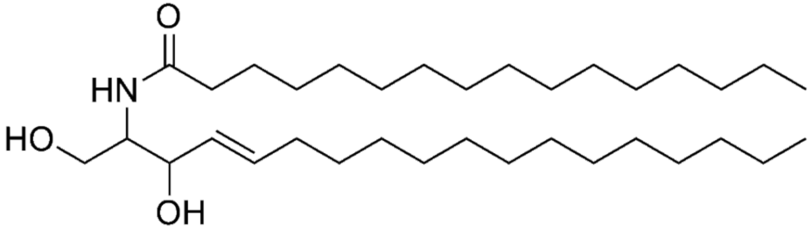


ascorbic acid



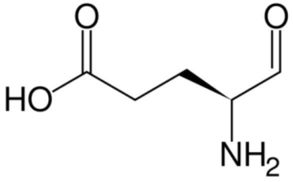
citric acid





Metabolites

C16-Ceramide



glutamate

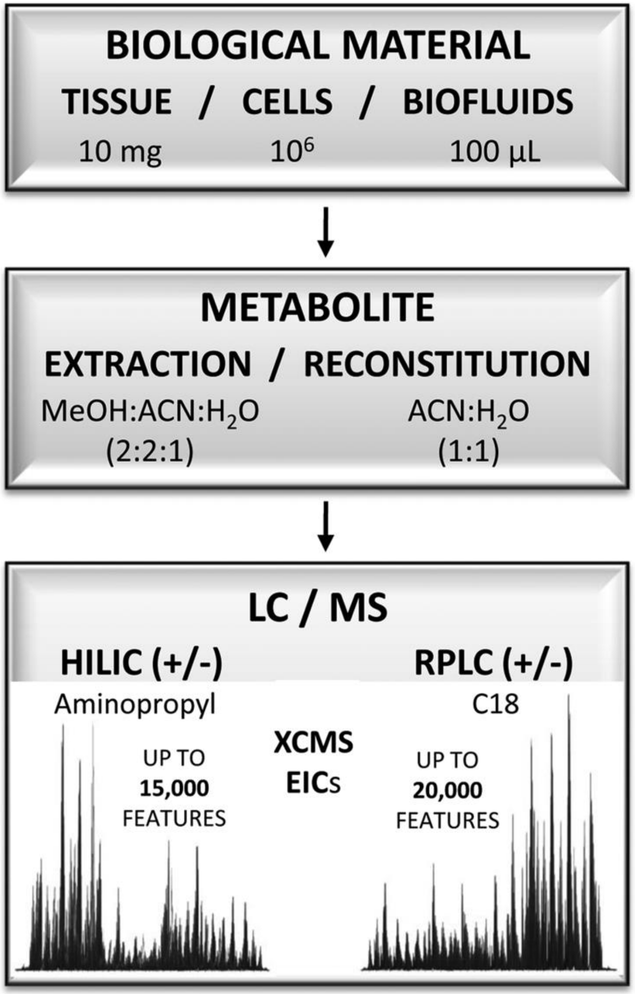
High chemical diversity

Yanes et al. 2011, Ivanisevic et al. 2013 AnalChem

Extraction and LC/MS

Experimental design

toward ‘omic scale metabolite profiling



BIOLOGICAL MATERIAL

TISSUE / CELLS / BIOFLUIDS

10 mg 10^6 100 µL

METABOLITE

EXTRACTION / RECONSTITUTION

MeOH:ACN:H_2O (2:2:1)

ACN:H_2O (1:1)

LC / MS

HILIC (+/-)
Aminopropyl
UP TO **15,000** FEATURES

XCMS
EICs

RPLC (+/-)
C18
UP TO **20,000** FEATURES

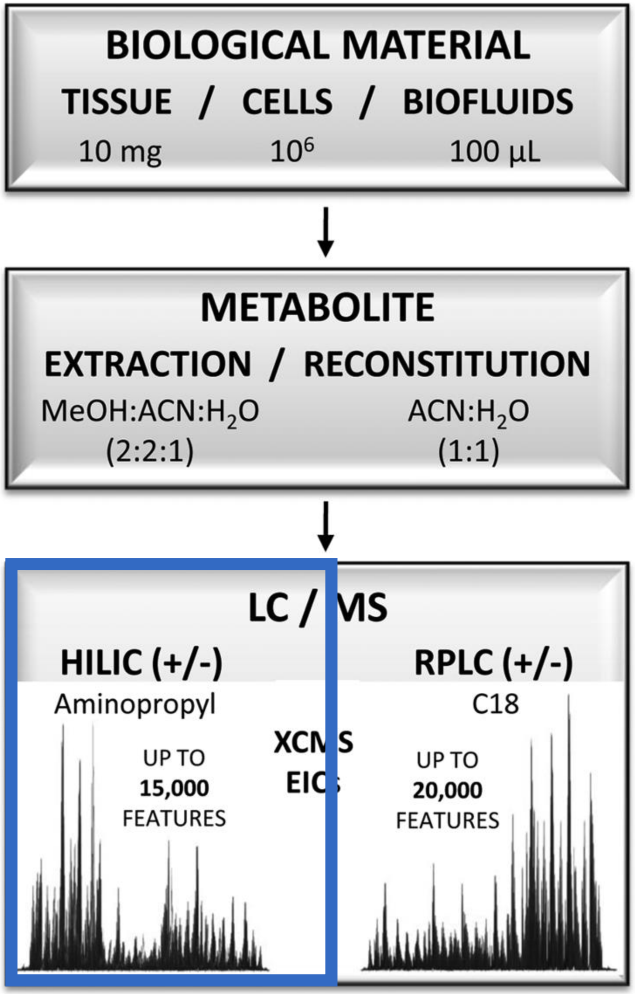
Profiling of both polar and non-polar metabolites with ONE single extraction

Ivanisevic, J., et al. Anal. Chem, 2014

Extraction and LC/MS

Experimental design

toward ‘omic scale metabolite profiling



BIOLOGICAL MATERIAL

TISSUE / CELLS / BIOFLUIDS

10 mg 10^6 100 µL

METABOLITE

EXTRACTION / RECONSTITUTION

MeOH:ACN:H_2O (2:2:1) ACN:H_2O (1:1)

LC / MS

HILIC (+/-) Aminopropyl

UP TO **15,000** FEATURES

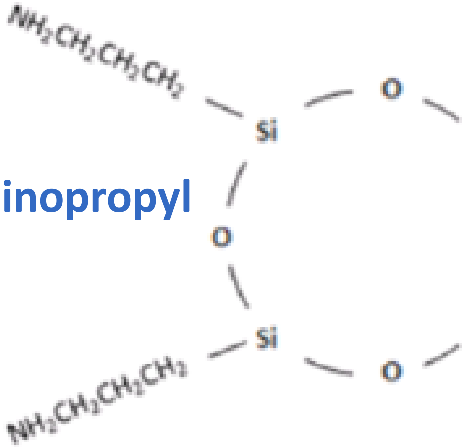
XCMS

EIC

RPLC (+/-) C18

UP TO **20,000** FEATURES

Profiling of both polar and non-polar metabolites with ONE single extraction



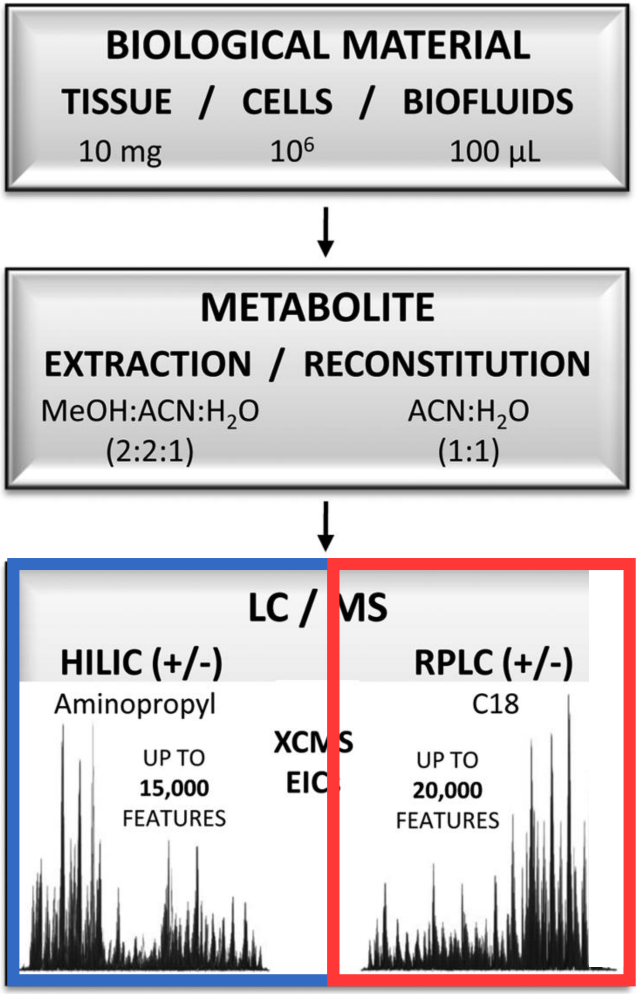
aminopropyl

Ivanisevic, J., et al. Anal. Chem, 2014

Extraction and LC/MS

Experimental design

toward 'omic scale metabolite profiling



BIOLOGICAL MATERIAL

TISSUE / CELLS / BIOFLUIDS

10 mg 10^6 100 µL

METABOLITE

EXTRACTION / RECONSTITUTION

$MeOH:ACN:H_2O$ (2:2:1) $ACN:H_2O$ (1:1)

LC / MS

HILIC (+/-)

Aminopropyl

UP TO **15,000** FEATURES

XCMS

EIC

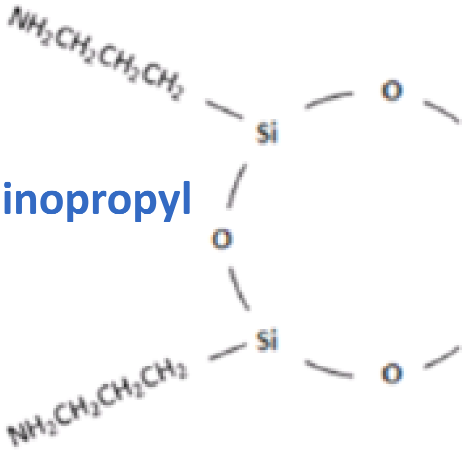
RPLC (+/-)

C18

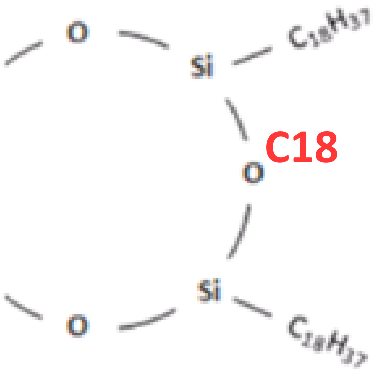
UP TO **20,000** FEATURES

Profiling of both polar and non-polar metabolites with ONE single extraction

aminopropyl



C18



Ivanisevic, J., et al. Anal. Chem, 2014

Extraction and LC/MS

Experimental design

toward ‘omic scale metabolite profiling

BIOLOGICAL MATERIAL

TISSUE / CELLS / BIOFLUIDS

10 mg 10^6 100 µL

↓

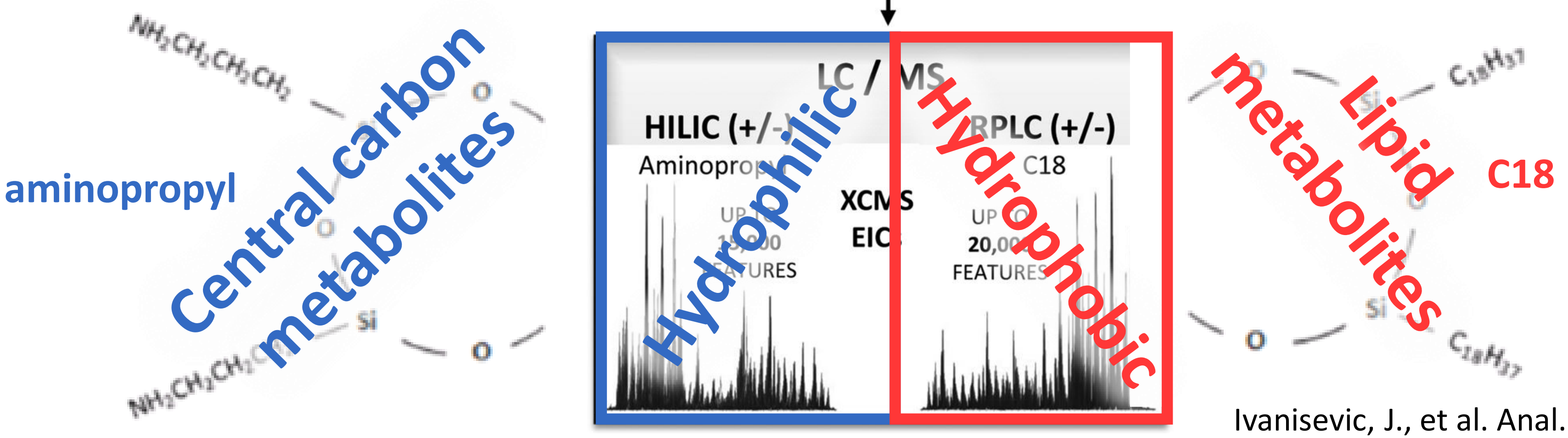
METABOLITE

EXTRACTION / RECONSTITUTION

MeOH:ACN:H_2O (2:2:1) ACN:H_2O (1:1)

↓

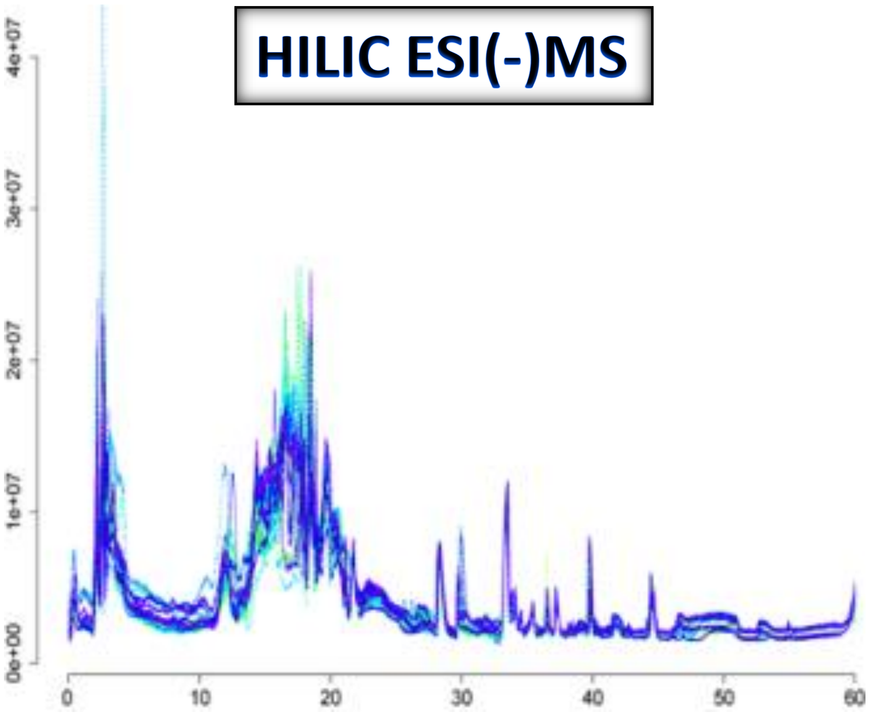
Profiling of both polar and non-polar metabolites with ONE single extraction



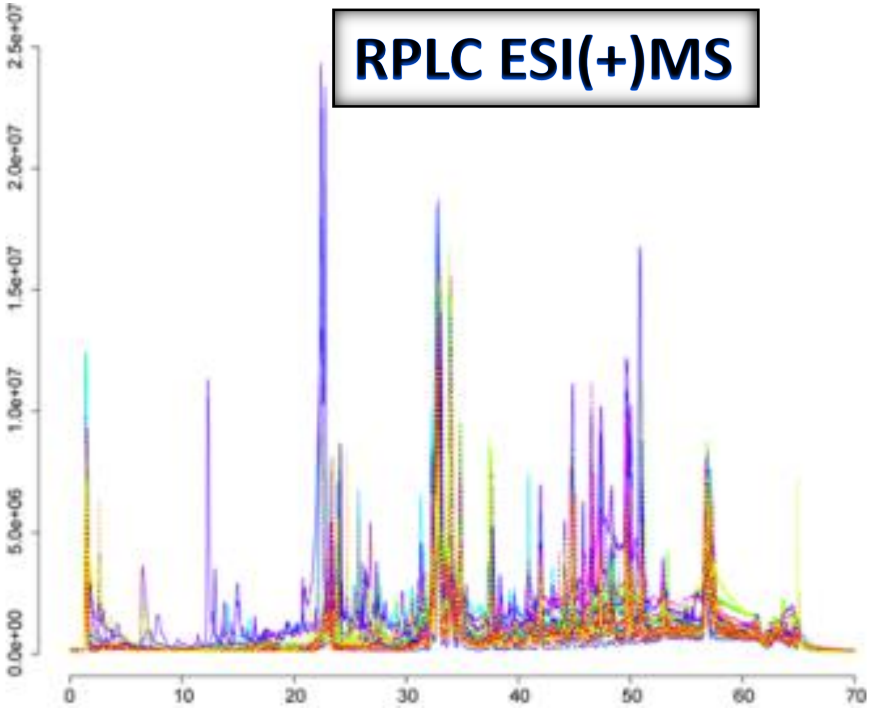
Ivanisevic, J., et al. Anal. Chem, 2014

Extraction and LC/MS

When time and sample are limited the highest number of unique, biologically relevant metabolites can be detected in:



HILIC ESI(-)MS



RPLC ESI(+)MS

Methods for extraction

Dependent on sample type

- -**non-mechanical methods** (Liquids: plasma, urine, cells, CSF)
  - -freeze/thawing to break open the cell
  - -addition of solvent

- -**mechanical method** (Solids: organs, tissues, feces)
  - -homogenizer

Non-Mechanical Extraction Example

Cell pellet extraction

5-10 million cells + ice cold water: acetonitrile: methanol (1:2:2 v/v/v)

Non-Mechanical Extraction Example

Cell pellet extraction

Vortex

Freeze

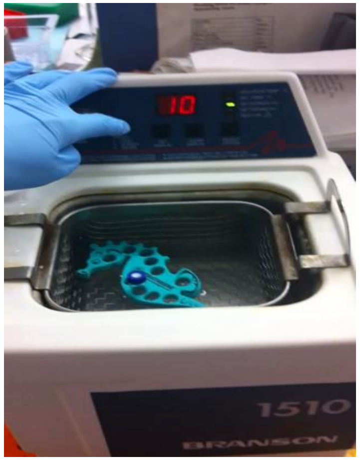
Thaw

Non-Mechanical Extraction Example

Cell pellet extraction

Transfer Samples

Sonicate



Repeat Twice

Store –20˚C Overnight/1 h

Non-Mechanical Extraction Example

Cell pellet extraction

Remove from –20˚C

Transfer to centrifuge tubes

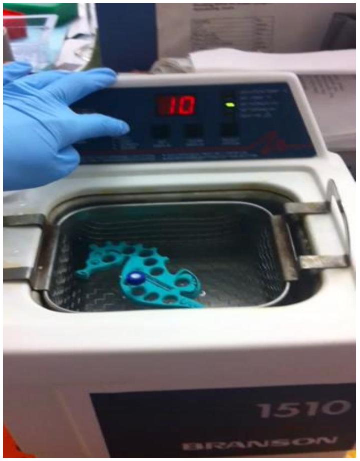
Evaporate solvent

Non-Mechanical Extraction Example

Cell pellet extraction

Resuspend
in LCMS
Solvent

Sonicate



Centrifuge and
extract supernatant

Sample is ready
for LC-MS

Cell Culture Specifics

Cells in suspension vs. adherent cell lines

- Avoid using trypsin

Cell Culture Specifics

Cells in suspension vs. adherent cell lines

- Avoid using trypsin

Without protein measurement

Cell Culture Specifics

Cells in suspension vs. adherent cell lines

- Avoid using trypsin

Without protein measurement

- Wash cells with PBS
- Add 1 mL ice cold solvent and scrape cells into a vial

Cell Culture Specifics

Cells in suspension vs. adherent cell lines

- Avoid using trypsin

Without protein measurement

- Wash cells with PBS
- Add 1 mL ice cold solvent and scrape cells into a vial
- Dry samples in a lyophilizer, extract same weight per sample, and reconstitute in 1 mg per 100 µL solvent.

Cell Culture Specifics

Cells in suspension vs. adherent cell lines

- Avoid using trypsin

With protein measurement

Cell Culture Specifics

Cells in suspension vs. adherent cell lines

- Avoid using trypsin

With protein measurement

- Wash cells with PBS
- Add 1 ml ice cold water on liquid nitrogen and scrape cells into a vial

Cell Culture Specifics

Cells in suspension vs. adherent cell lines

- Avoid using trypsin

With protein measurement

- Wash cells with PBS
- Add 1 ml ice cold water on liquid nitrogen and scrape cells into a vial
- Carry out freeze-thawing steps and collect aliquot for protein assay
- Then add solvent

Biofluid Specifics

Plasma/Urine/CSF

- Add 400 µL of cold solvent to 100 µL plasma to maintain ratio of 2:2:1 (MeOH:ACN:H2O)

- Vortex and sonicate 10 min

Methods for extraction

Dependent on sample type

- **non-mechanical methods** (Liquids: plasma, urine, cells, CSF)
  - freeze/thawing to break open the cell
  - addition of solvent

- **mechanical method** (Solids: organs, tissues, feces)
  - homogenizer

Mechanical Extraction Example

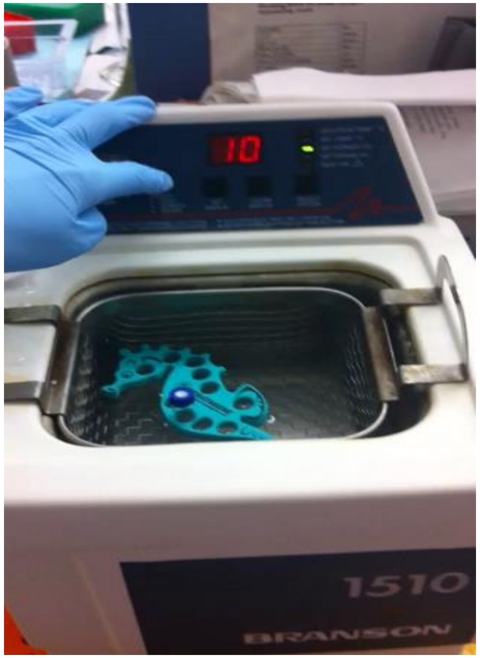
Tissue extraction

10 mg tissue +
Ice cold water: methanol (1:4 v/v) +
Homogenization Beads → 1mm Zirconia beads
1mm Glass beads

Mechanical Extraction Example

Tissue extraction

Homogenizer



Sonicate

Store –20˚C
Overnight/1 h

Mechanical Extraction Example

Tissue extraction

Remove from
–20˚C

Transfer to centrifuge tubes

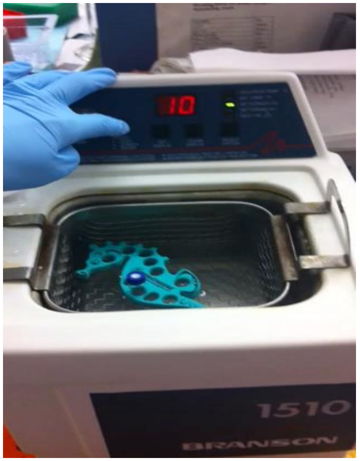
Evaporate solvent

Mechanical Extraction Example

Tissue extraction

Resuspend in LCMS Solvent

Sonicate



Centrifuge and extract supernatant

Sample is ready for LC-MS

Normalization Strategies

Tissues – Weight

– Protein

Normalization Strategies

Tissues – Weight
– Protein

Cells – Cell number
– Protein or DNA concentration/estimation

Normalization Strategies

Tissues – Weight

– Protein

Cells – Cell number

– Protein or DNA concentration/estimation

Biofluids – Volume for most biofluids

– Challenging to normalize urine

Quality Sample Prep

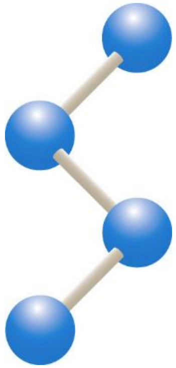
Reduce background ions: salts/buffers/surfactants, PEG, solvents (DMSO), polymers/plasticizers

Reduce biological contamination: keratin, cross contamination, cellular media

Reduce sample degradation: properly store samples after preparation

Quality sample prep → Higher quality results

May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Statistical Analysis***

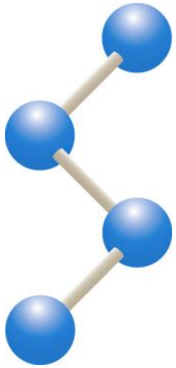
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

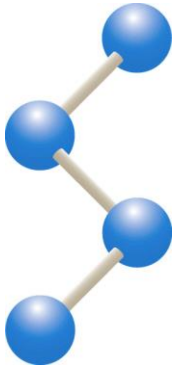
May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

May 16th
---- 12:30 pm Begin ----
---- 3:00 pm Break ----
---- 04:30 pm Finish ----



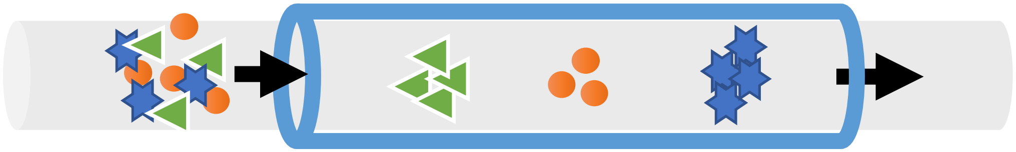
Fundamental Metabolomics

May 16th

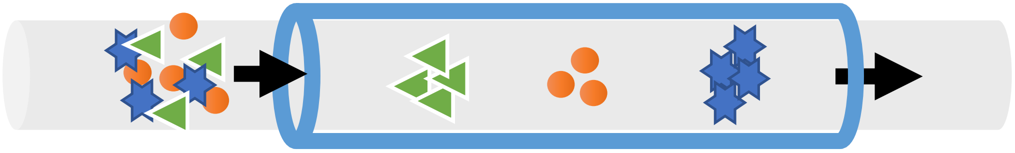
- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

May 16th
---- 12:30 pm Begin ----
---- 3:00 pm Break ----
---- 04:30 pm Finish ----

Chromatography



Chromatography



Sample

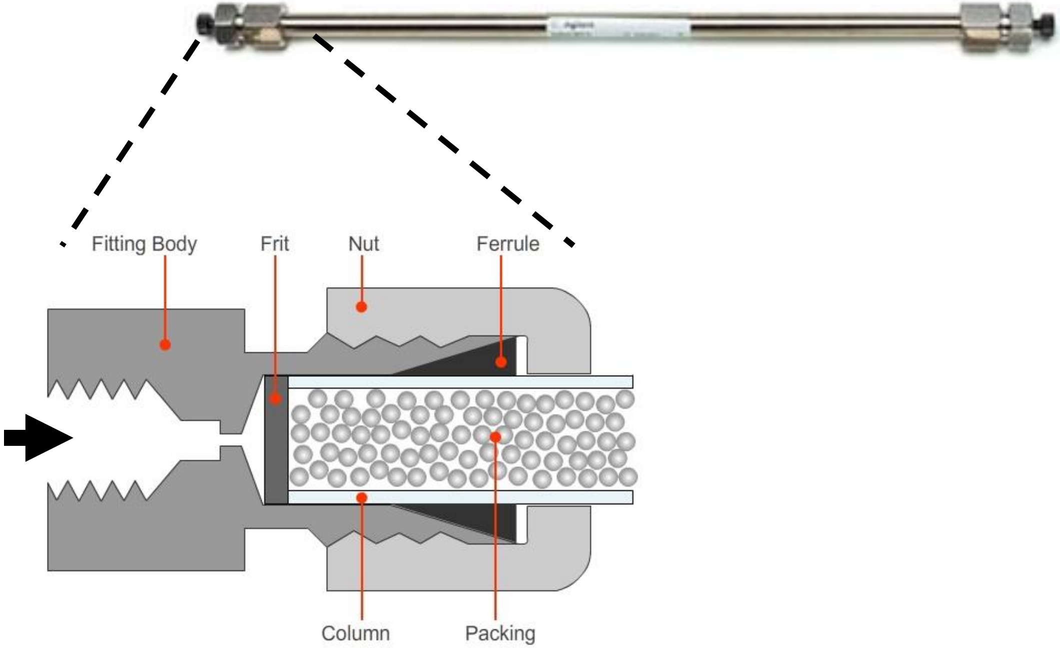
Detector
(MS)

Mobile Phase
(MP)

Interaction

Stationary Phase
(SP)

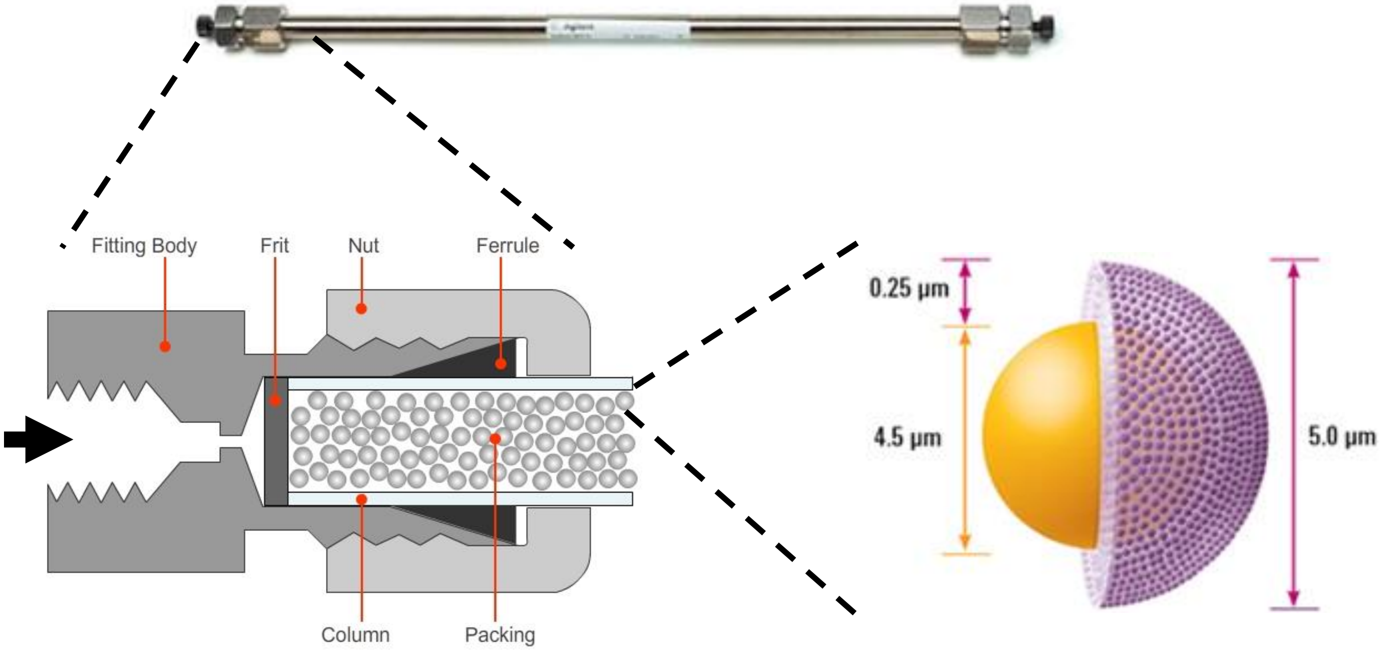
Chromatography



Sample MP

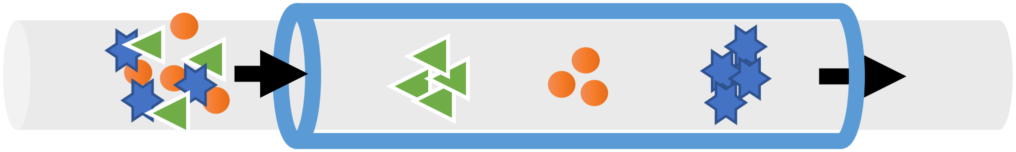
Chromatography

Sample MP



HPLC C18

Chromatography



High affinity to stationary phase

Low affinity to stationary phase

Detector (MS)

Sample

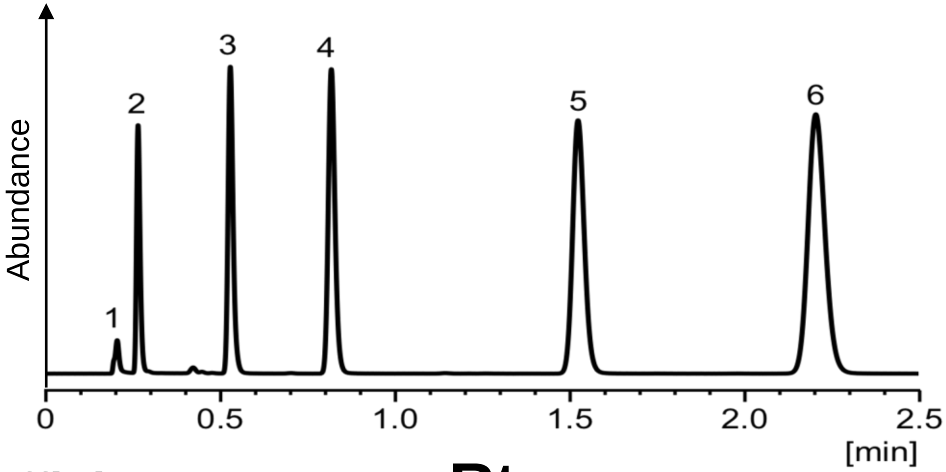
Mobile Phase

↕

Interaction

Stationary Phase

Chromatography



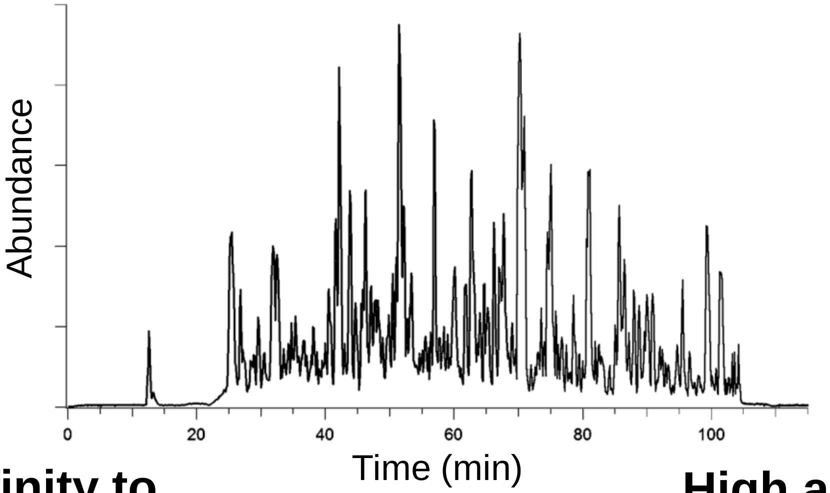
Targeted Analysis

Low affinity to stationary phase

Rt

High affinity to stationary phase

Chromatography



Untargeted Analysis

Low affinity to stationary phase

High affinity to stationary phase

Why chromatography in metabolomics?

1. Reduce ion suppression for MS

Better sensitivity and increase in detected molecules

Why chromatography in metabolomics?

1. Reduce ion suppression for MS

Better sensitivity and increase in detected molecules

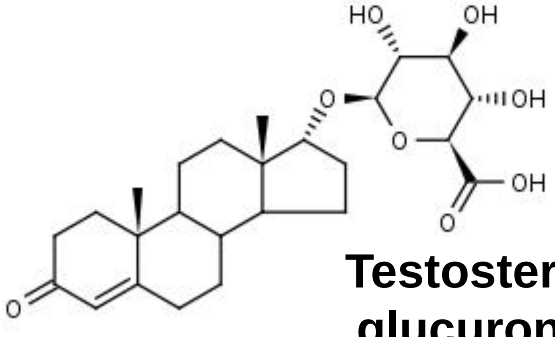
2. Separate isobaric molecules

Why chromatography in metabolomics?

1. Reduce ion suppression for MS

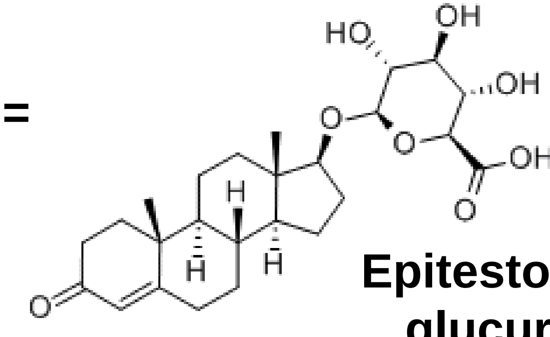
Better sensitivity and increase in detected molecules

2. Separate isobaric molecules



Testosterone glucuronide

= *m/z* =



Epitestosterone glucuronide

Why chromatography in metabolomics?

1. Reduce ion suppression for MS

Better sensitivity and increase in detected molecules

2. Separate isobaric molecules
3. Retention time for metabolite ID

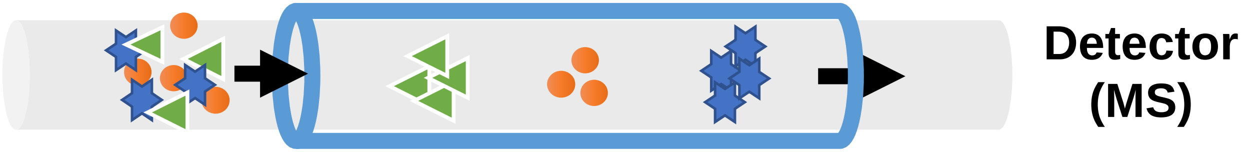
Why chromatography in metabolomics?

1. Reduce ion suppression for MS

Better sensitivity and increase in detected molecules

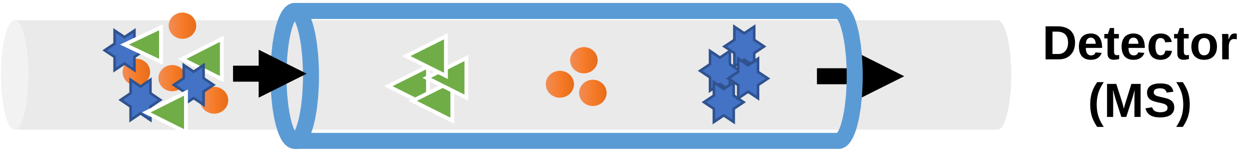
2. Separate isobaric molecules
3. Retention time for metabolite ID
4. High-quality MS/MS data

Factors affecting Chromatography



- Mobile phase (MP) (composition, pH, salts content, elution gradient)
- Temperature
- Type of stationary phase (SP)

Factors affecting Chromatography



- Mobile phase (MP) (composition, pH, salts content, elution gradient)
- Temperature
- Type of stationary phase (SP)

MF Composition

In water

Organic

%A + %B = MF

- **+ salts** for buffer effect
  (**pH** and **ionic strength** control)

MF Composition

In water

Organic

%A + %B = MF

- **+ salts** for buffer effect (**pH** and **ionic strength** control)

Reversed-phase

ESI + A: 100% H_20 + 0.1% FA

B: 100% ACN + 0.1% FA

ESI- A: 100% H_20 + 1mM NH_4F

B: 100% ACN

HILIC

ESI- A: 95% H_20 + 20mM NH_4Ac + 20mM NH_4OH

B: 95% ACN

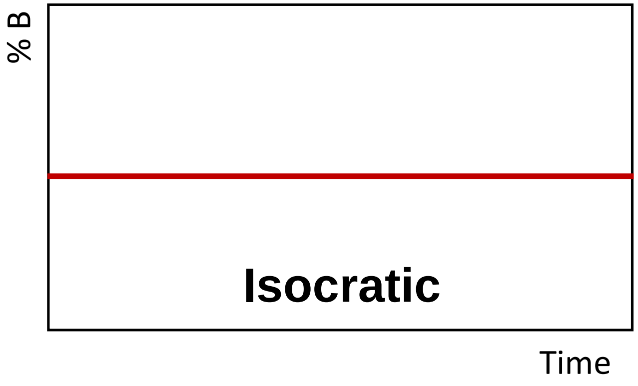
MF Composition

In water

Organic

%A + %B = MF

- **+ salts** for buffer effect (**pH** and **ionic strength** control)
- Isocratic elution



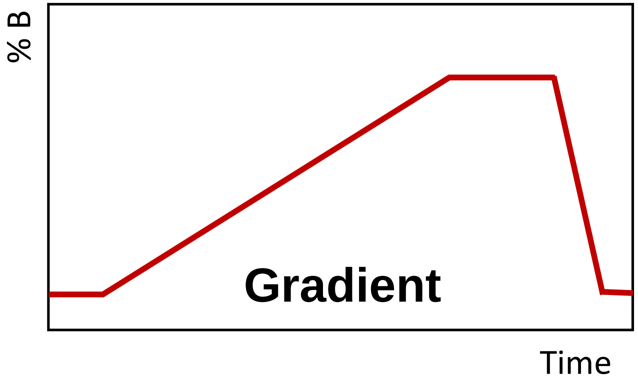
MF Composition

In water

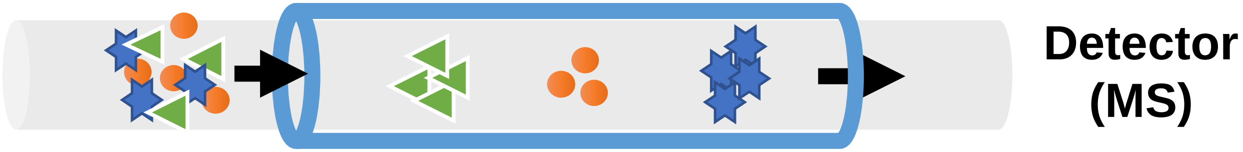
Organic

%A + %B = MF

- **+ salts** for buffer effect (**pH** and **ionic strength** control)
- Isocratic elution
- Gradient elution



Factors affecting Chromatography



- Mobile phase (MP) (composition, pH, salts content, elution gradient)
- Temperature
- Type of stationary phase (SP)

Column Temperature

T = 20 – 50˚C

- **Usually 27˚C**
- **T > 27 ˚C for the elution of highly retained compounds** (e.g. lipids in RP)

Column Temperature

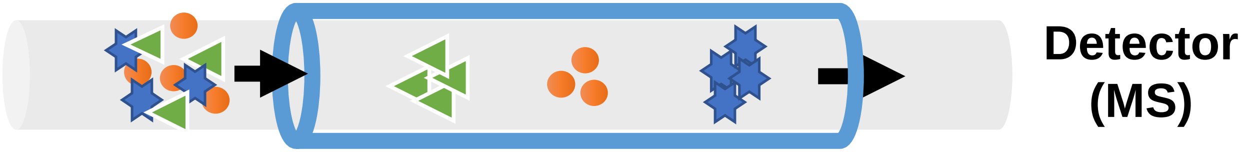
T = 20 – 50˚C

- **Usually 27˚C**
- **T > 27 ˚C for the elution of highly retained compounds** (e.g. lipids in RP)

Increase of T:

- <u>Increase elution</u> of highly retained compounds
- Allows higher <u>flow</u> rate for faster gradients
- Caution, can affect <u>metabolite stability</u> and <u>column lifetime</u>!

Factors Affecting Chromatography



- Mobile phase (MP) (composition, pH, salts content, elution gradient)
- Temperature
- Type of stationary phase (SP)

How do we choose the appropriate type of LC?

- **Time** & sample **number**

  -High-throughput?

How do we choose the appropriate type of LC?

- **Time** & sample **number**
  -High-throughput?

- **Type of metabolites** you wish to see?
  - Polar vs. non-polar
  - Specific metabolite class
  - Isomers

Speed: HPLC vs. UPLC

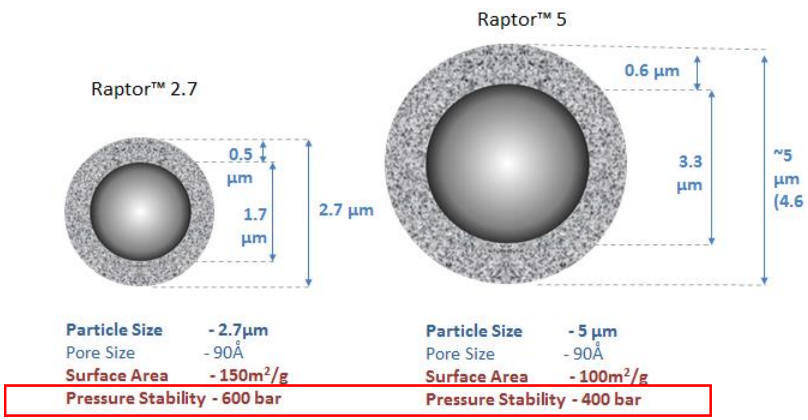
- **Time** & sample **number**

  - High-throughput?

A) **High** Performance Liquid Chromatography (**HPLC**)
B) **Ultra** Performance Liquid Chromatography (**UPLC**)

Speed: HPLC vs. UPLC

A) **High** Performance Liquid Chromatography (HPLC)
B) **Ultra** Performance Liquid Chromatography (UPLC)



UPLC **HPLC**

Particle size

Pressure

UPLC

- Operates at higher pressure
- Smaller particles size

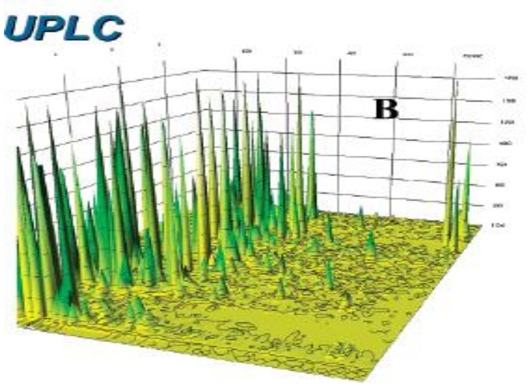
Speed: HPLC vs. UPLC

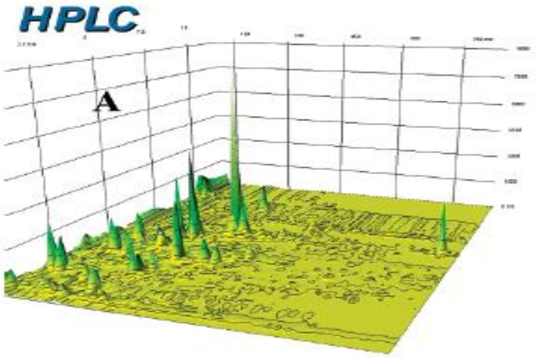
UPLC

- Improved resolution
- Generally higher sensitivity
- Higher pressure stability
- Higher mass spec scan speed required
- **Shorter gradients**

HPLC

- Lower likelihood of column clogging
- Lower back pressure
- Can use mass spec with slower scan speed





Speed: LC vs. NO LC ?

- Flow Injection Analysis (FIA) – 1-2min/sample
- No chromatography → <u>cannot separate isobars</u>, <u>high ion suppression</u> and <u>lower sensitivity</u>.
- Useful for:
  - **real-time** metabolomic profiling (Link, H., Nat. Met. 2015)
  - **high-throughput** studies > 1000 samples

How do we choose the appropriate type of LC?

- **Time** & sample **number**
  -High-throughput?

- **Type of metabolites** you want to detect?
  - Polar vs. non-polar
  - Specific metabolite class
  - Isomers

LC Choices in Metabolomics

- **HPLC (High Performance Liquid Chromatography)**
  - Reversed Phase
  - Hydrophilic Interaction Liquid Chromatography: HILIC

- **UPLC (Ultra Performance Liquid Chromatography)**
  - Reversed Phase
  - Hydrophilic Interaction Liquid Chromatography: HILIC

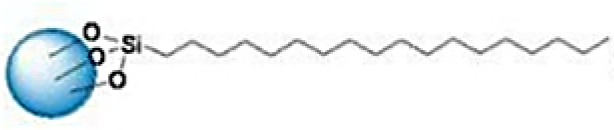
LC Choices in Metabolomics

- **HPLC (High Performance Liquid Chromatography)**
  - **Reversed Phase**
  - Hydrophilic Interaction Liquid Chromatography: HILIC

- **UPLC (Ultra Performance Liquid Chromatography)**
  - **Reversed Phase**
  - Hydrophilic Interaction Liquid Chromatography: HILIC

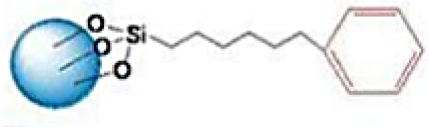
Reversed-Phase

C18



C8

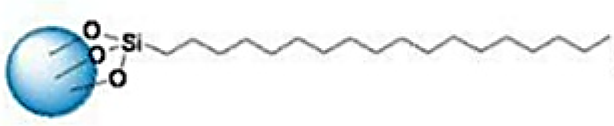
Phenyl



Reversed-Phase

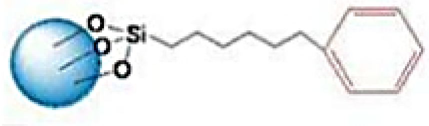
Hydrophobic Interactions

C18



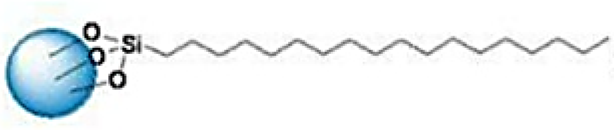
C8

Phenyl



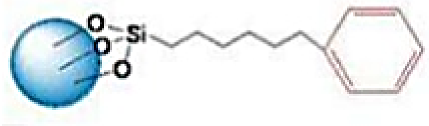
Reversed-Phase

C18



C8

Phenyl

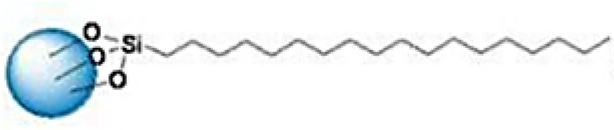


Hydrophobic Interactions

Non-polar Metabolites

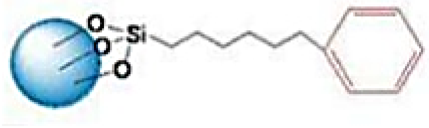
Reversed-Phase

C18



C8

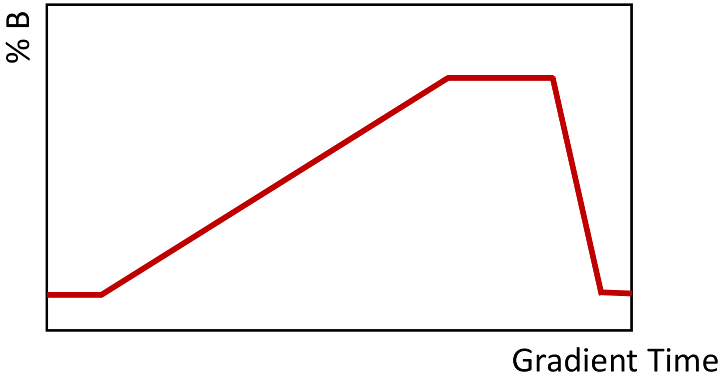
Phenyl



Hydrophobic Interactions

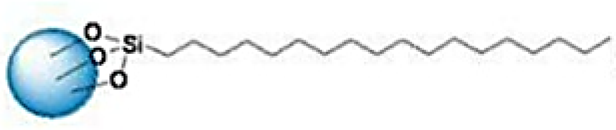
Non-polar Metabolites

Elution Gradient: From High Water to High Organic Solvents (ACN or IPA)



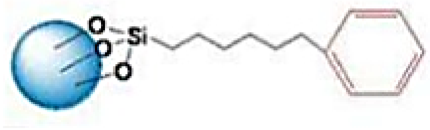
Reversed-Phase

C18



C8

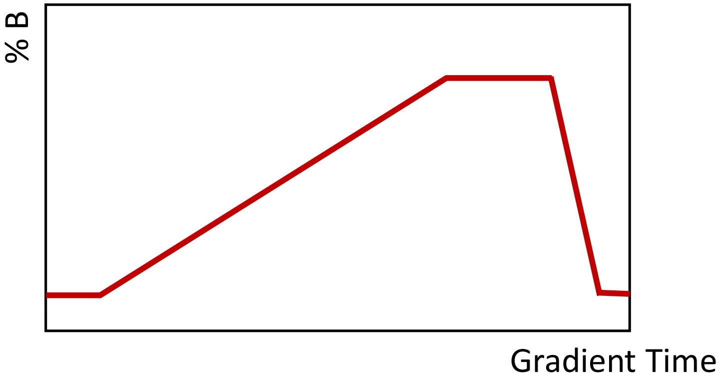
Phenyl



Hydrophobic Interactions

Non-polar Metabolites

Elution Gradient: From High Water to High Organic Solvents (ACN or IPA)



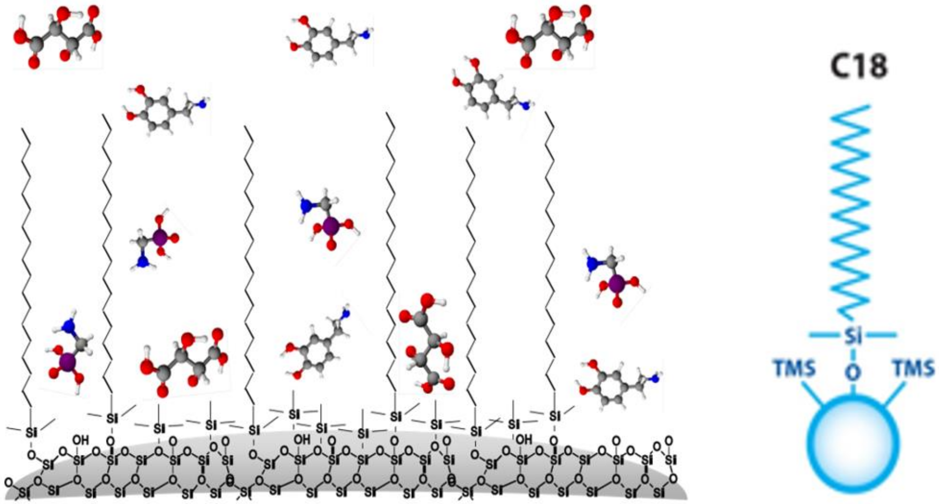
Mobile Phase Modifier:

- 0.1% Formic Acid (ESI+)

- 1 mM NH_4F (ESI-)

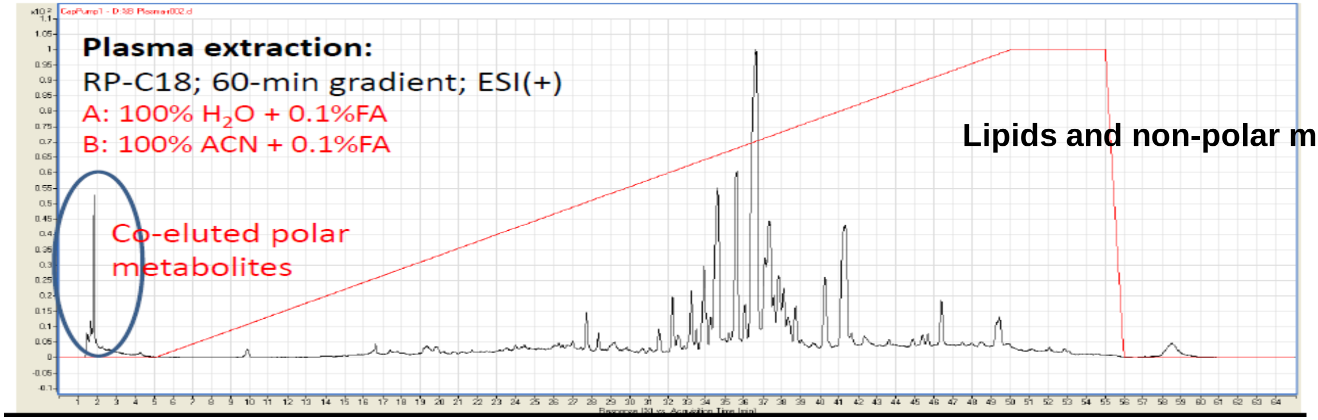
Reversed-Phase

- Large number of metabolomic studies use **C18 RP LC**



Reversed-Phase

- Large number of metabolomic studies use **C18 RP LC**



Plasma extraction:

RP-C18; 60-min gradient; ESI(+)

A: 100% H_2O + 0.1%FA

B: 100% ACN + 0.1%FA

Co-eluted polar metabolites

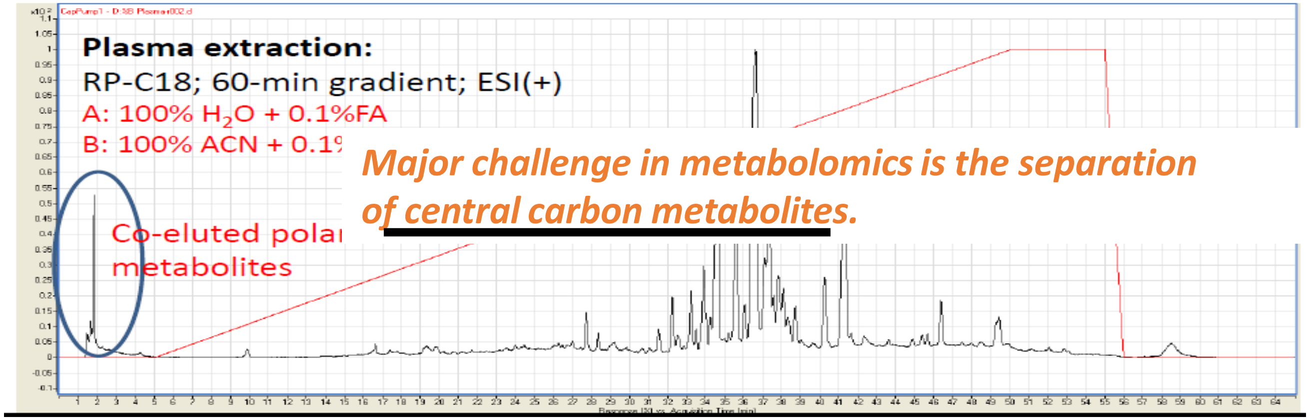
Lipids and non-polar metabolites

Reversed-Phase

- Large number of metabolomic studies use **C18 RP LC**

- **Highly polar molecules are not retained**

__void volume__: glycolysis,
pentose-phosphate, TCA,…
(lots of important structural isomers)



Plasma extraction:
RP-C18; 60-min gradient; ESI(+)
A: 100% H_2O + 0.1%FA
B: 100% ACN + 0.1%

Co-eluted polar metabolites

Major challenge in metabolomics is the separation of central carbon metabolites.

LC Choices in Metabolomics

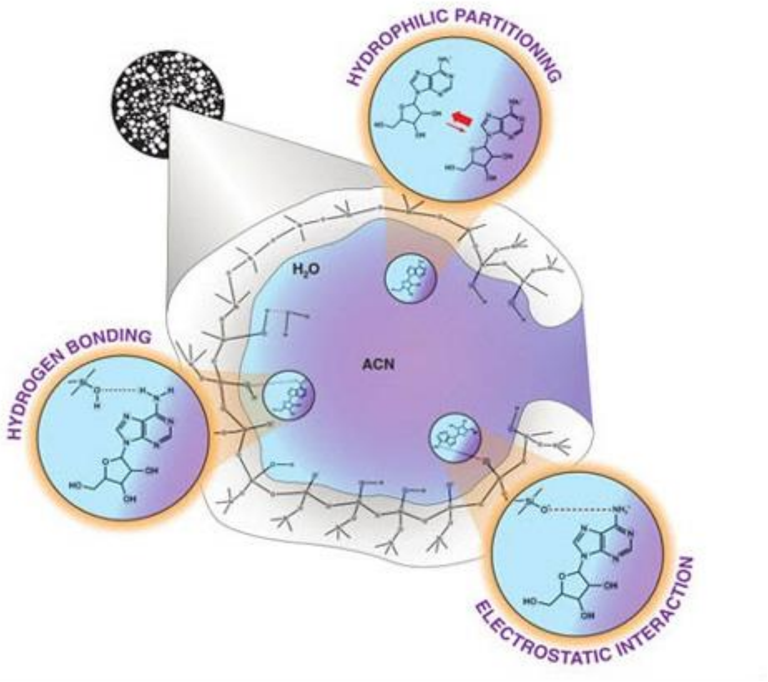
- **HPLC (High Performance Liquid Chromatography)**
  - Reversed Phase
  - **Hydrophilic Interaction Liquid Chromatography: HILIC**

- **UPLC (Ultra Performance Liquid Chromatography)**
  - Reversed Phase
  - **Hydrophilic Interaction Liquid Chromatography: HILIC**

Hydrophilic Interaction Liquid Chromatography (HILIC)

- **Derivatization** of silica stationary phase enhances retention of polar molecules **by hydrogen bonding or ionic interactions**.

- Elution gradient: from high organic with low % water and salts, to high water and salts.



- MF forms a **water-rich layer** on the surface of the polar stationary phase, creating a **liquid/liquid extraction system.**

- The retention is based on the **hydrophilic partitioning of the analytes into the water-enriched SP** and **electrostatic interactions with the charged SP**.

Hydrophilic Interaction Liquid Chromatography (HILIC)

- **Derivatization** of silica stationary phase enhances retention of polar molecules **by hydrogen bonding or ionic interactions**.

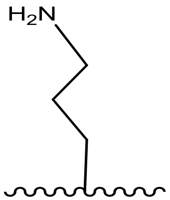
- Elution gradient: from high organic to high water.

Silica, $Si\text{-}OH \rightleftharpoons Si\text{-}O^{(-)}H^{(+)}$

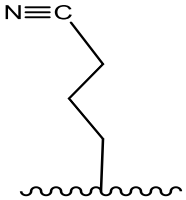
Amine, $-(CH_2)_3\text{-}NH_2$

Diol, $-(CH_2)_3\text{-}O\text{-}CH_2\text{-}(CHOH)CH_2OH$

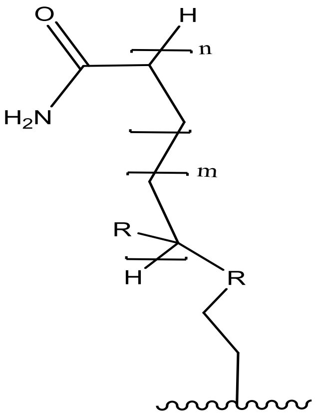
Amide, $-(CH_2)_n\text{-}(CO)NH_2$



amino



cyano



Amide

Hydrophilic Interaction Liquid Chromatography (HILIC)

- Increasingly popular, particularly for targeted analysis of metabolites in **central carbon metabolism.**

Hydrophilic Interaction Liquid Chromatography (HILIC)

- Increasingly popular, particularly for targeted analysis of metabolites in **central carbon metabolism.**

- Highly hydrophobic compounds are not retained.

Hydrophilic Interaction Liquid Chromatography (HILIC)

- Increasingly popular, particularly for targeted analysis of metabolites in **central carbon metabolism.**

- Highly hydrophobic compounds are not retained.

- Long equilibration times and shorter column lifetime.

Hydrophilic Interaction Liquid Chromatography (HILIC)

- Increasingly popular, particularly for targeted analysis of metabolites in **central carbon metabolism.**

- Highly hydrophobic compounds are not retained.

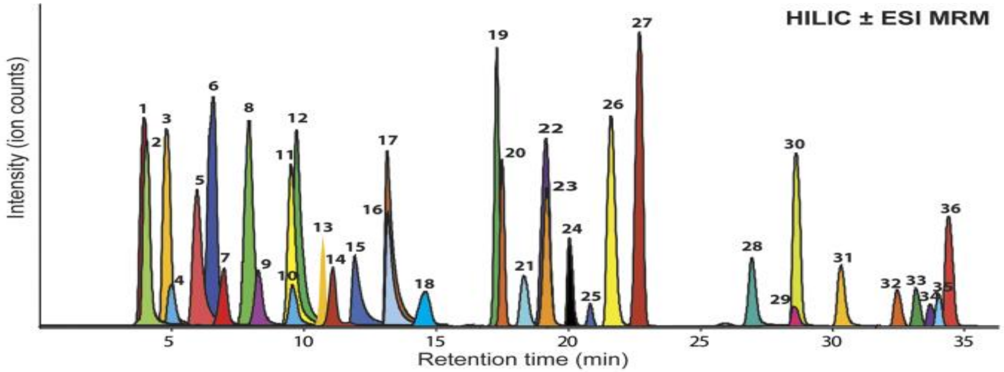
- Long equilibration times and shorter column lifetime.

- At high water/salt/pH buffer, SP can degrade rapidly.

HILIC for Untargeted Metabolomics

HILIC COLUMN for untargeted metabolomics:

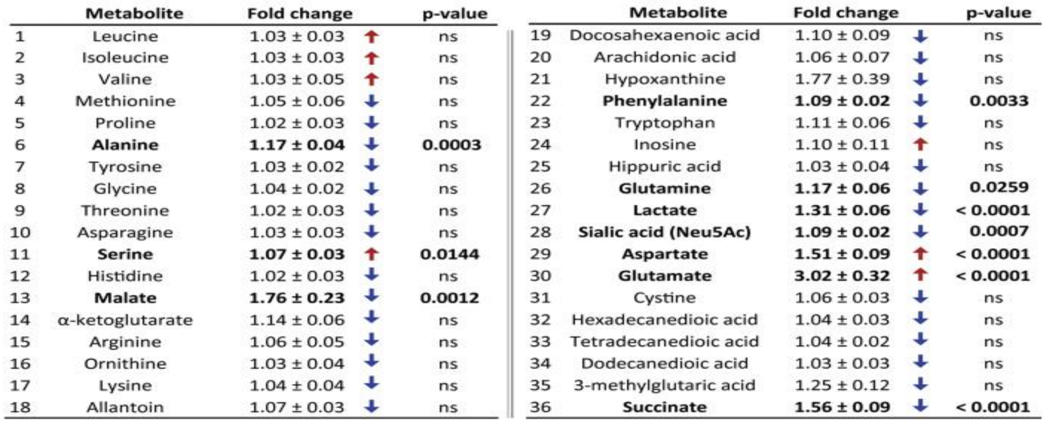
- HPLC: Luna NH_2 (3 µm, 150 x 1 mm) column Phenomenex (Ivanisevic, J, Anal. Chem. 85(14), 2013)



Mobile Phase (pH: 9.7)

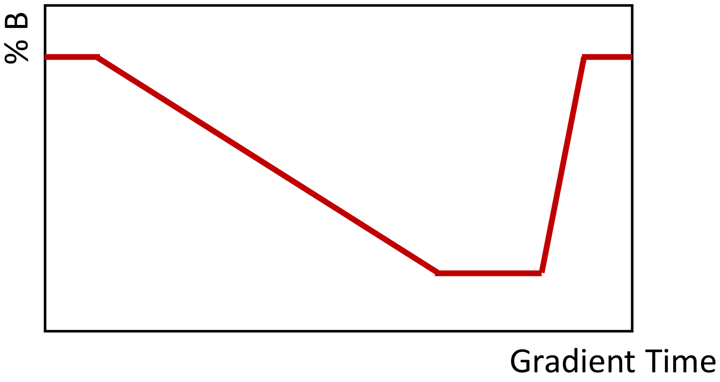
A: 95% H20 +20mM NH_4Ac
+20mM NH_4OH

B: 95% ACN



	Metabolite	Fold change		p-value
1	Leucine	1.03 ± 0.03	↑	ns
2	Isoleucine	1.03 ± 0.03	↑	ns
3	Valine	1.03 ± 0.05	↑	ns
4	Methionine	1.05 ± 0.06	↓	ns
5	Proline	1.02 ± 0.03	↓	ns
6	**Alanine**	**1.17 ± 0.04**	↓	**0.0003**
7	Tyrosine	1.03 ± 0.02	↓	ns
8	Glycine	1.04 ± 0.02	↓	ns
9	Threonine	1.02 ± 0.03	↓	ns
10	Asparagine	1.03 ± 0.03	↓	ns
11	**Serine**	**1.07 ± 0.03**	↑	**0.0144**
12	Histidine	1.02 ± 0.03	↓	ns
13	**Malate**	**1.76 ± 0.23**	↓	**0.0012**
14	α-ketoglutarate	1.14 ± 0.06	↓	ns
15	Arginine	1.06 ± 0.05	↓	ns
16	Ornithine	1.03 ± 0.04	↓	ns
17	Lysine	1.04 ± 0.04	↓	ns
18	Allantoin	1.07 ± 0.03	↓	ns
19	Docosahexaenoic acid	1.10 ± 0.09	↓	ns
20	Arachidonic acid	1.06 ± 0.07	↓	ns
21	Hypoxanthine	1.77 ± 0.39	↓	ns
22	**Phenylalanine**	**1.09 ± 0.02**	↓	**0.0033**
23	Tryptophan	1.11 ± 0.06	↓	ns
24	Inosine	1.10 ± 0.11	↑	ns
25	Hippuric acid	1.03 ± 0.04	↓	ns
26	**Glutamine**	**1.17 ± 0.06**	↓	**0.0259**
27	**Lactate**	**1.31 ± 0.06**	↓	**< 0.0001**
28	**Sialic acid (Neu5Ac)**	**1.09 ± 0.02**	↓	**0.0007**
29	**Aspartate**	**1.51 ± 0.09**	↑	**< 0.0001**
30	**Glutamate**	**3.02 ± 0.32**	↑	**< 0.0001**
31	Cystine	1.06 ± 0.03	↓	ns
32	Hexadecanedioic acid	1.04 ± 0.03	↓	ns
33	Tetradecanedioic acid	1.04 ± 0.02	↓	ns
34	Dodecanedioic acid	1.03 ± 0.03	↓	ns
35	3-methylglutaric acid	1.25 ± 0.12	↓	ns
36	**Succinate**	**1.56 ± 0.09**	↓	**< 0.0001**

Elution Gradient: From High Organic to High Water

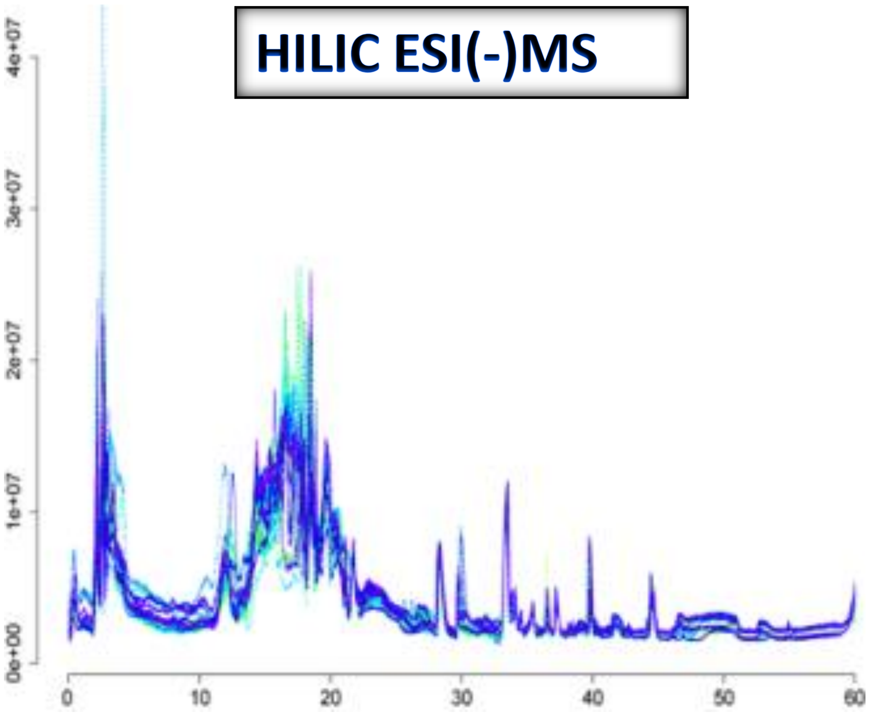


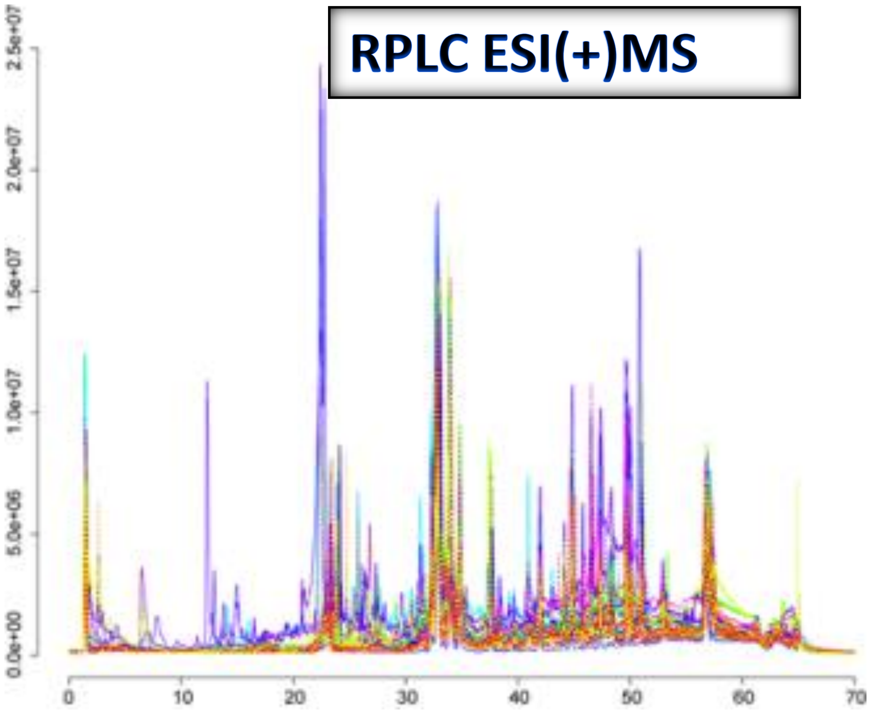
Dual Separation In Metabolomics

When ***time and sample are limited*** *the highest number of unique, biologically relevant metabolites can be detected in:*

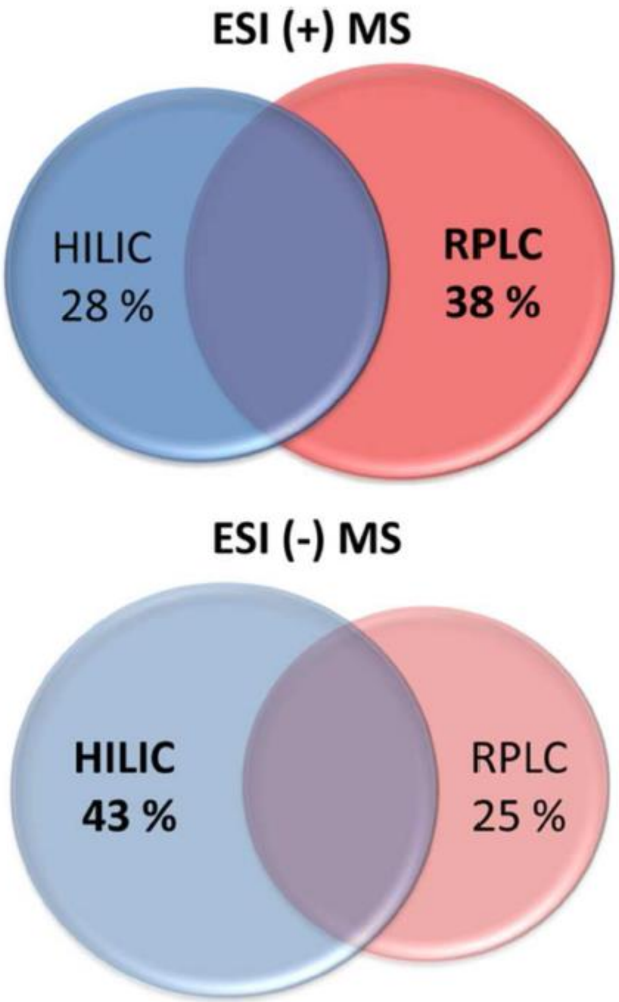
Dual Separation In Metabolomics

*When **time and sample are limited** the highest number of unique, biologically relevant metabolites can be detected in:*





Dual Separation In Metabolomics



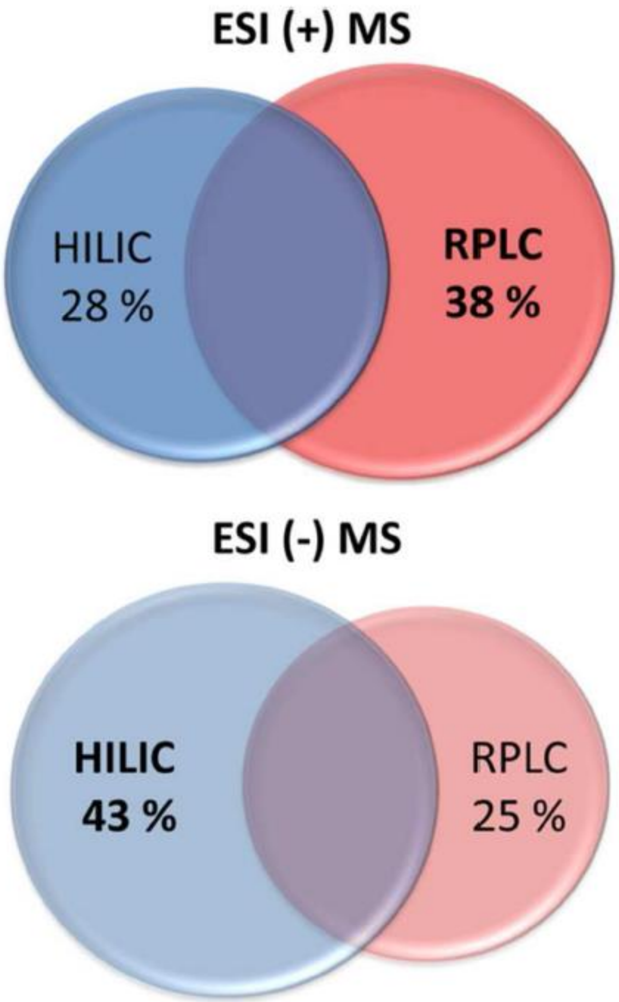
Good coverage:

RPLC ESI(+) AND **HILIC ESI (-)**

(Ivanisevic, J, Anal. Chem. 85(14), 2013)

LC: Take-Home Message

Type of Stationary Phase



Mobile Phase Composition

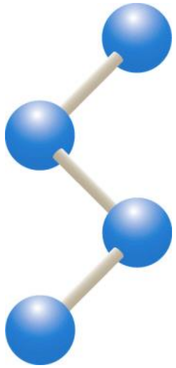
pH

Ionic Strength

Temperature

Time Factor

HPLC vs. UPLC



Fundamental Metabolomics

May 16th

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

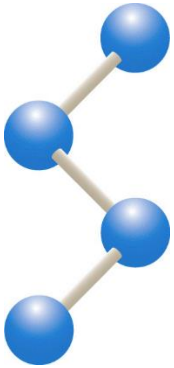
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Statistical Analysis***

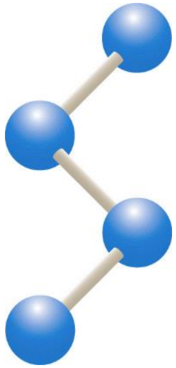
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Statistical Analysis***

May 16th

---- 12:30 pm Begin ----

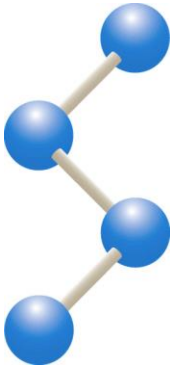
---- 3:00 pm Break ----

---- 04:30 pm Finish ----

Fundamental Metabolomics

Nuclear magnetic resonance (NMR)

- Good for absolute quantification
- Not sensitive enough for broad metabolome coverage



Fundamental Metabolomics

Nuclear magnetic resonance (NMR)

❑ Good for absolute quantification
❑ Not sensitive enough for broad metabolome coverage

Gas chromatography-mass spectrometry (GC-MS)

❑ Good chromatographic separation
❑ Not suitable for thermo-labile metabolites

Fundamental Metabolomics

Nuclear magnetic resonance (NMR)

- ☐ Good for absolute quantification
- ☐ Not sensitive enough for broad metabolome coverage

Gas chromatography-mass spectrometry (GC-MS)

- ☐ Good chromatographic separation
- ☐ Not suitable for thermo-labile metabolites

✓ Liquid chromatography-mass spectrometry (LC-MS)

- ✓ Good sensitivity and wide metabolome coverage

Fundamental Metabolomics

LC-MS based Untargeted Metabolomics

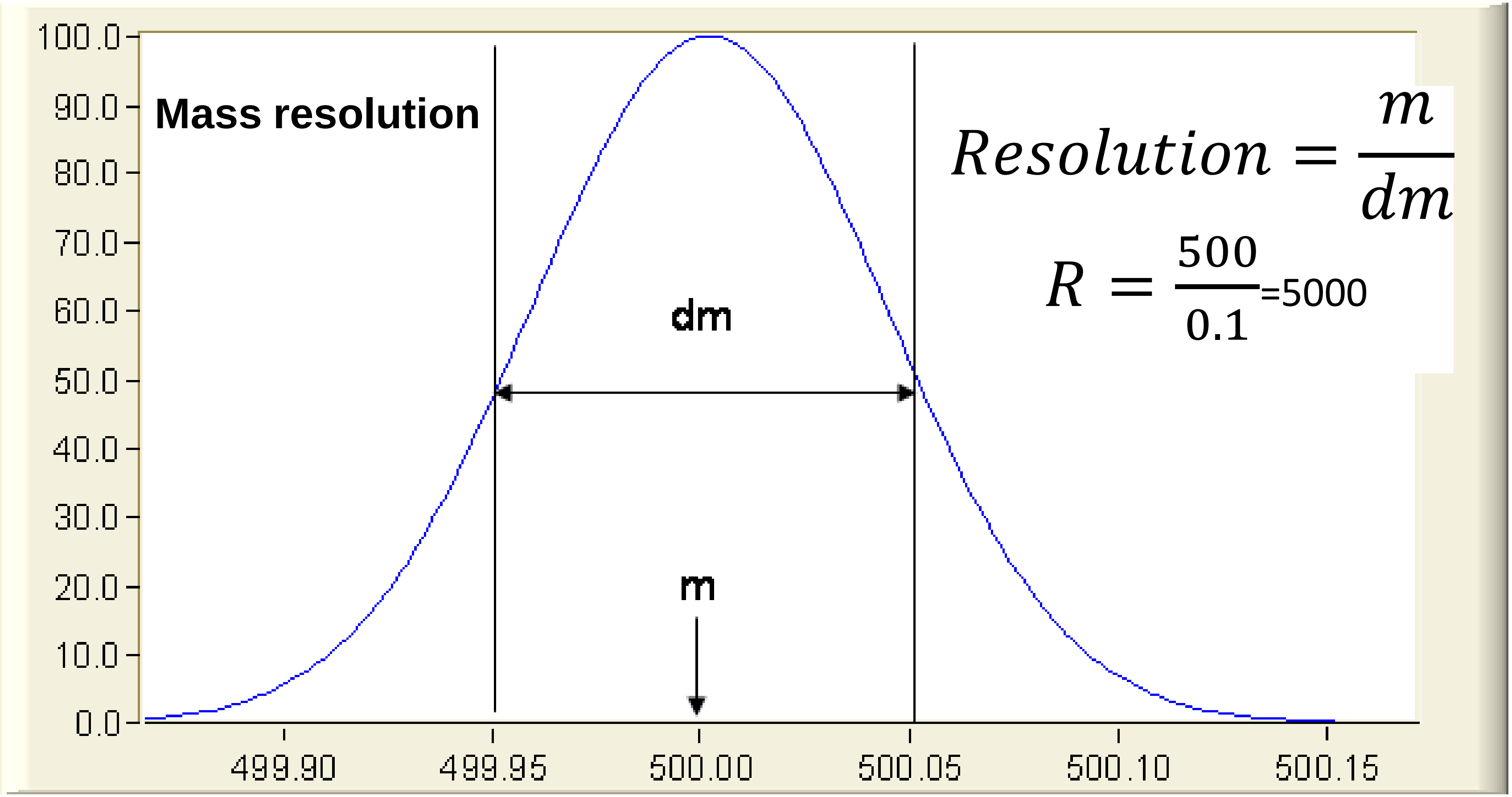
- **Choice of MS**
- LC-MS combination
- Analytical sequence
- Data processing
- Metabolite identification

Fundamental Metabolomics

Choice of MS

Mass resolution
Mass accuracy

Fundamental Metabolomics



Mass resolution

$$Resolution = \frac{m}{dm}$$

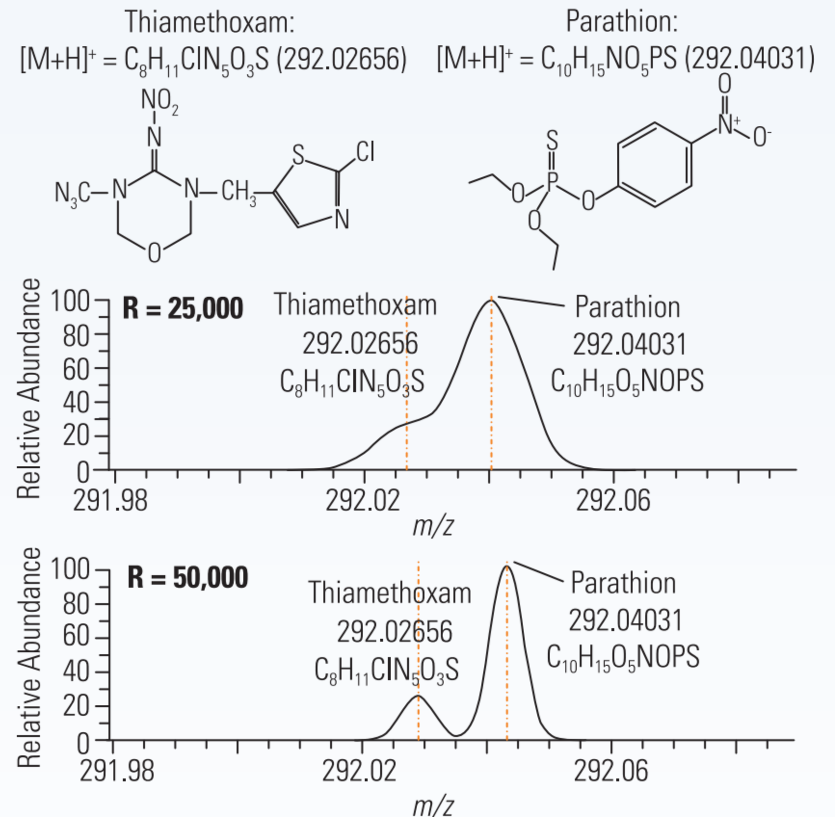
$$R = \frac{500}{0.1} = 5000$$

Fundamental Metabolomics

How Much Resolution Is Required

Fundamental Metabolomics

- The higher the better, there may never be enough resolution



How Much Resolution Is Required

- The higher the better, there may never be enough resolution
- Typical metabolomic profiling requires mass resolution higher than 15,000

Fundamental Metabolomics

Mass Accuracy

Fundamental Metabolomics

Mass Accuracy

- Mass error = (theoretical *m/z*) – (measured *m/z*)

Fundamental Metabolomics

Mass Accuracy

- Mass error = (theoretical *m/z*) – (measured *m/z*)

- Mass error in parts per million (ppm) =

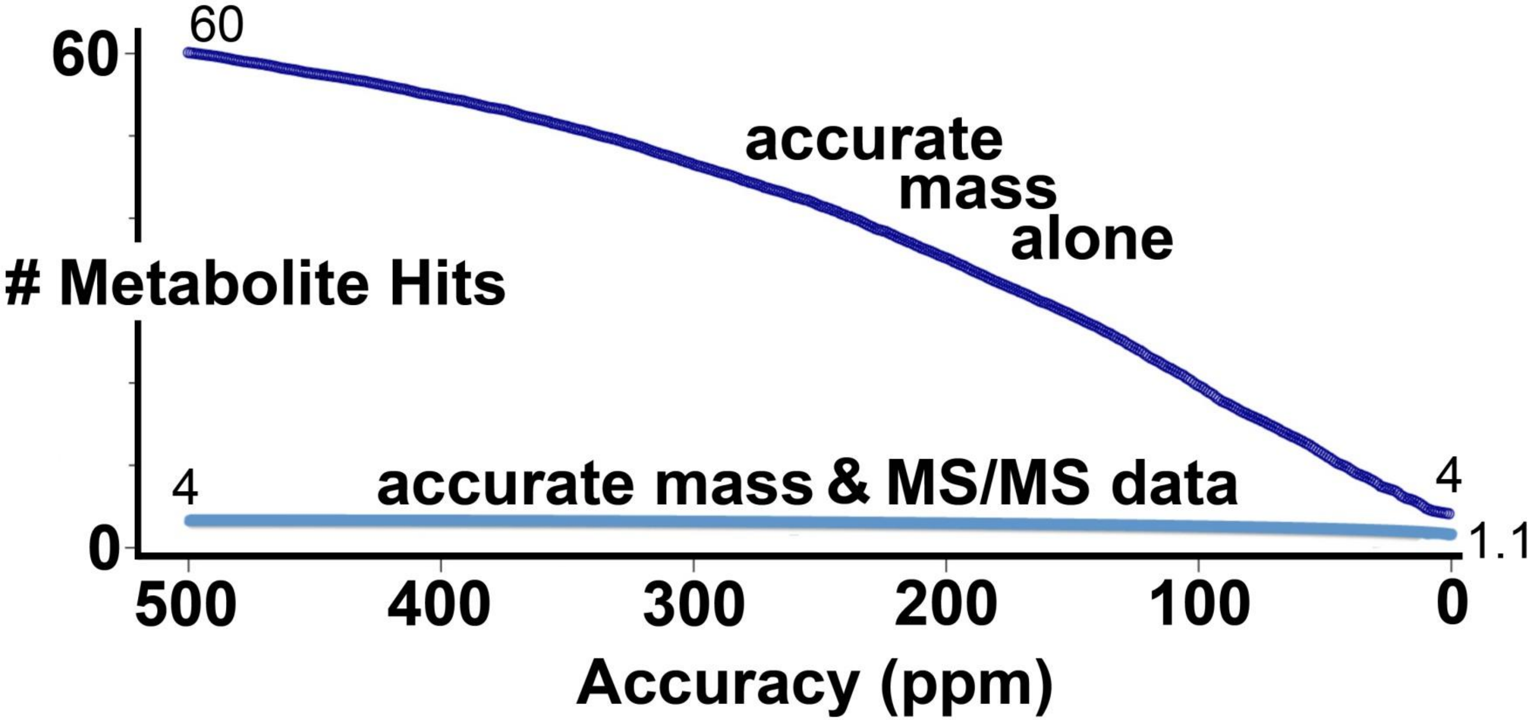
$X\ 10^6$

Fundamental Metabolomics

- Accurate mass is important for metabolite identification

Fundamental Metabolomics

- Accurate mass is important for metabolite identification



Fundamental Metabolomics

- Accurate mass is important for metabolite identification

- Accurate mass measurement can be used to determine the elemental composition

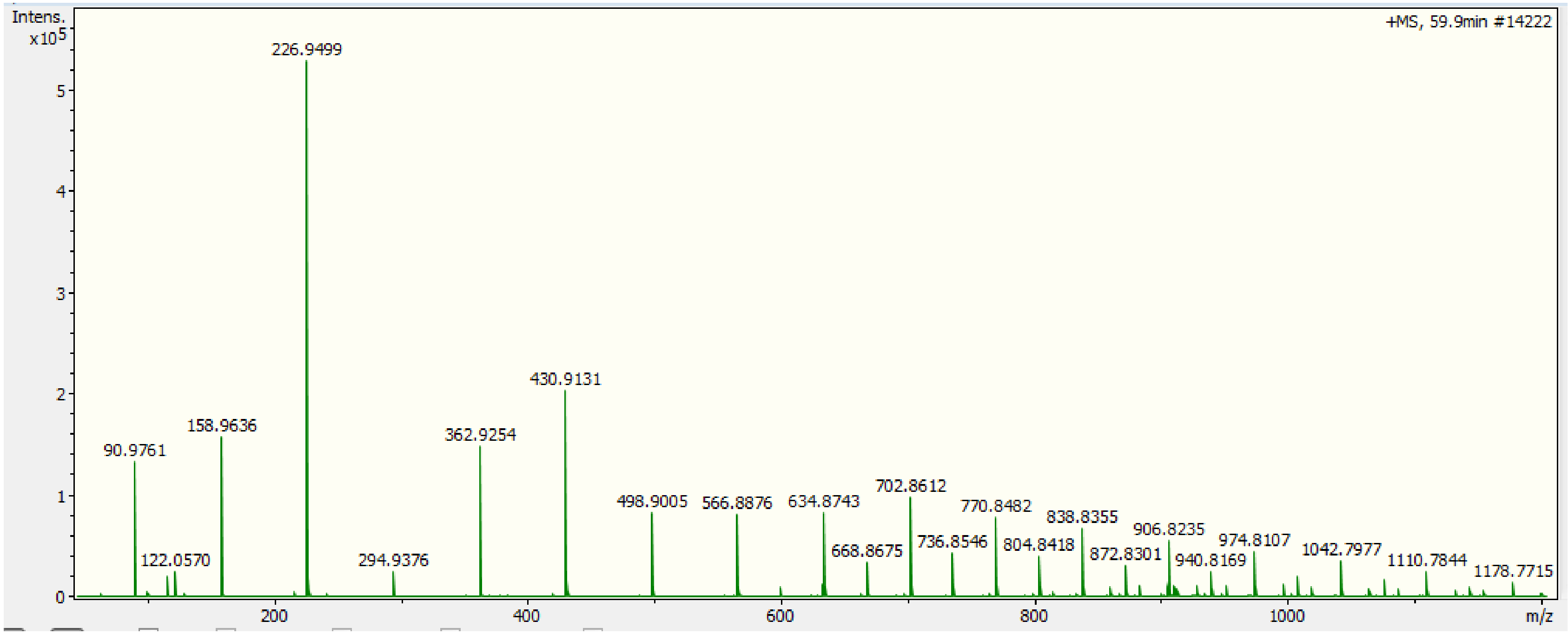
Fundamental Metabolomics

- Accurate mass is important for metabolite identification
- Accurate mass measurement can be used to determine the elemental composition
- Typical MS profiling requires mass accuracy better than 10 ppm

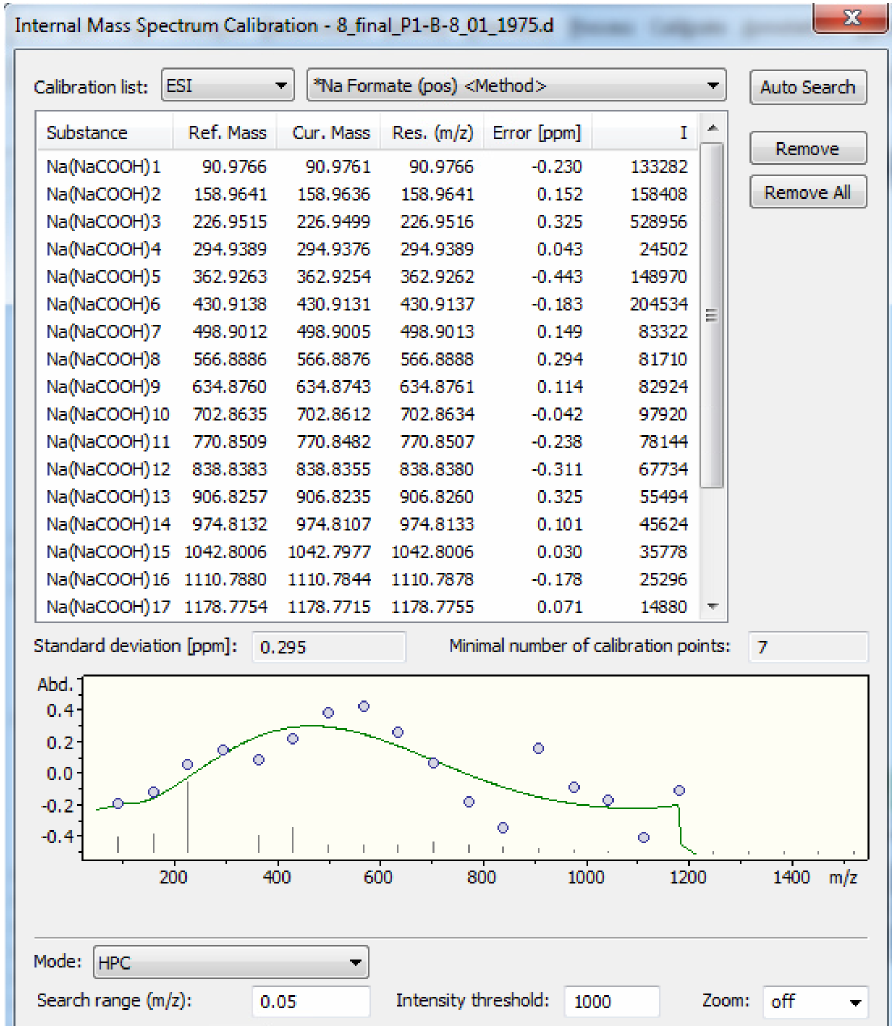
Fundamental Metabolomics

Achieve good mass accuracy through mass calibration

250 mM NaFA



Fundamental Metabolomics



Internal Mass Spectrum Calibration - 8_final_P1-B-8_01_1975.d

Calibration list: ESI

*Na Formate (pos) <Method>

Substance	Ref. Mass	Cur. Mass	Res. (m/z)	Error [ppm]	I
Na(NaCOOH)1	90.9766	90.9761	90.9766	-0.230	133282
Na(NaCOOH)2	158.9641	158.9636	158.9641	0.152	158408
Na(NaCOOH)3	226.9515	226.9499	226.9516	0.325	528956
Na(NaCOOH)4	294.9389	294.9376	294.9389	0.043	24502
Na(NaCOOH)5	362.9263	362.9254	362.9262	-0.443	148970
Na(NaCOOH)6	430.9138	430.9131	430.9137	-0.183	204534
Na(NaCOOH)7	498.9012	498.9005	498.9013	0.149	83322
Na(NaCOOH)8	566.8886	566.8876	566.8888	0.294	81710
Na(NaCOOH)9	634.8760	634.8743	634.8761	0.114	82924
Na(NaCOOH)10	702.8635	702.8612	702.8634	-0.042	97920
Na(NaCOOH)11	770.8509	770.8482	770.8507	-0.238	78144
Na(NaCOOH)12	838.8383	838.8355	838.8380	-0.311	67734
Na(NaCOOH)13	906.8257	906.8235	906.8260	0.325	55494
Na(NaCOOH)14	974.8132	974.8107	974.8133	0.101	45624
Na(NaCOOH)15	1042.8006	1042.7977	1042.8006	0.030	35778
Na(NaCOOH)16	1110.7880	1110.7844	1110.7878	-0.178	25296
Na(NaCOOH)17	1178.7754	1178.7715	1178.7755	0.071	14880

Standard deviation [ppm]: 0.295

Minimal number of calibration points: 7

Mode: HPC

Search range (m/z): 0.05

Intensity threshold: 1000

Zoom: off

Fundamental Metabolomics

Q-TOF

- Resolution: > 20,000
- Mass accuracy: < 1 ppm
- Mass range: up to 20,000 m/z
- Dynamic range: five orders of magnitude with femtogram-level sensitivity

Fundamental Metabolomics

Orbitrap

- Resolution: up to 140,000
- Mass accuracy: < 1 ppm
- Mass range: up to 6,000 m/z
- Dynamic range: four orders of magnitude with femtogram-level sensitivity

Fundamental Metabolomics

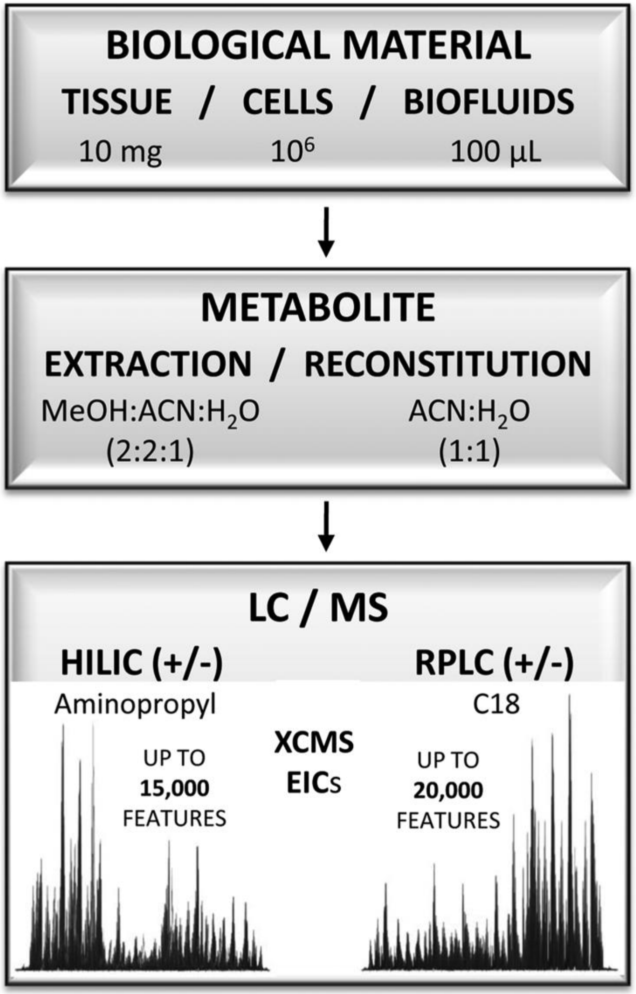
LC-MS based Untargeted metabolomics

- Choice of MS
- **LC-MS combination**
- Analytical sequence
- Data processing
- Metabolite identification

Fundamental Metabolomics

Toward 'omic scale metabolite profiling

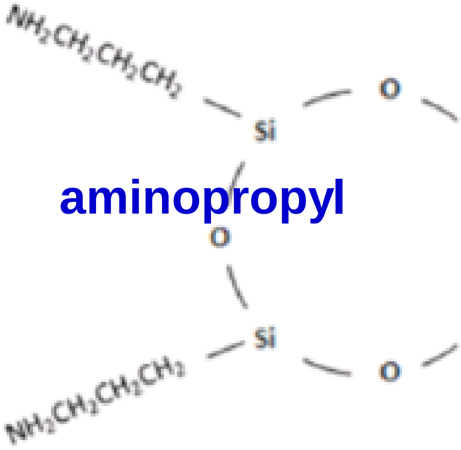
Profiling of both polar and non-polar metabolites with ONE single extraction



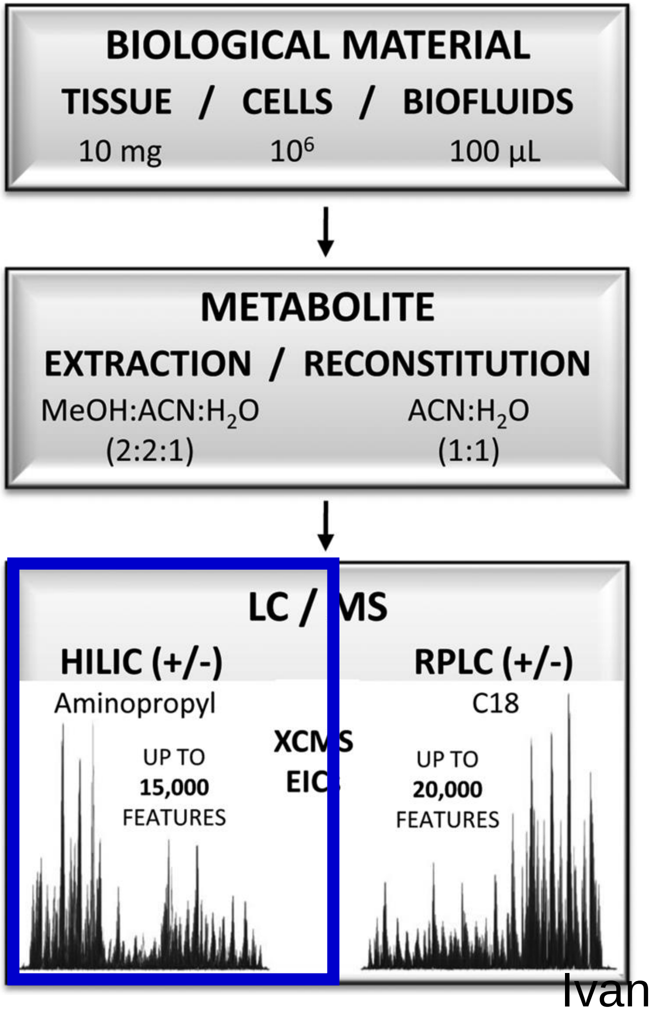
Ivanisevic, J., et al. Anal. Chem, 2014

Fundamental Metabolomics

Basic pH



aminopropyl



BIOLOGICAL MATERIAL

TISSUE / CELLS / BIOFLUIDS

10 mg | 10^6 | 100 µL

METABOLITE

EXTRACTION / RECONSTITUTION

MeOH:ACN:H_2O (2:2:1) | ACN:H_2O (1:1)

LC / MS

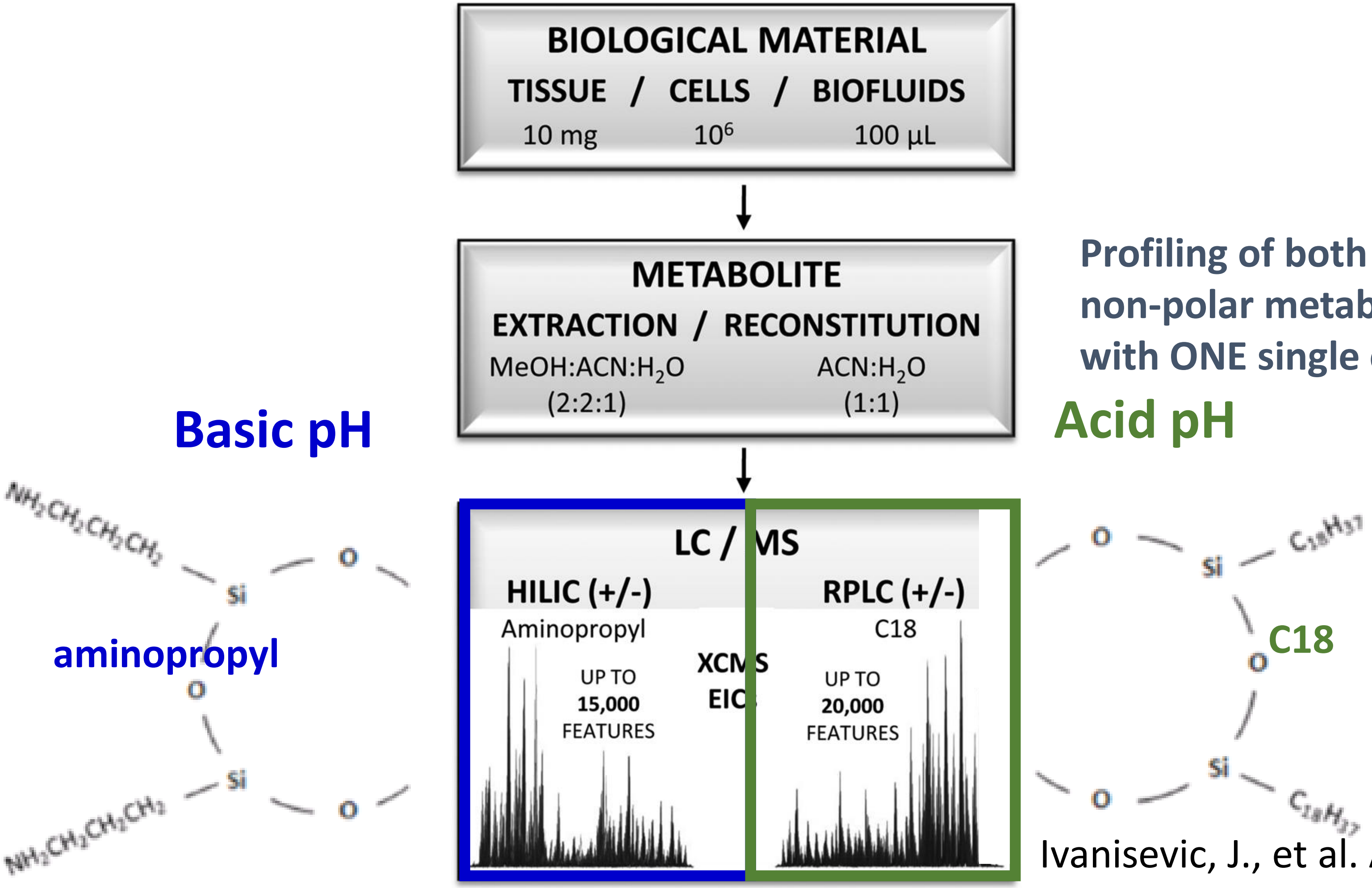
HILIC (+/-)
Aminopropyl
UP TO **15,000** FEATURES

XCMS
EIC

RPLC (+/-)
C18
UP TO **20,000** FEATURES

Ivanisevic, J., et al. Anal. Chem, 2

Fundamental Metabolomics



Profiling of both polar and non-polar metabolites with ONE single extraction

Basic pH

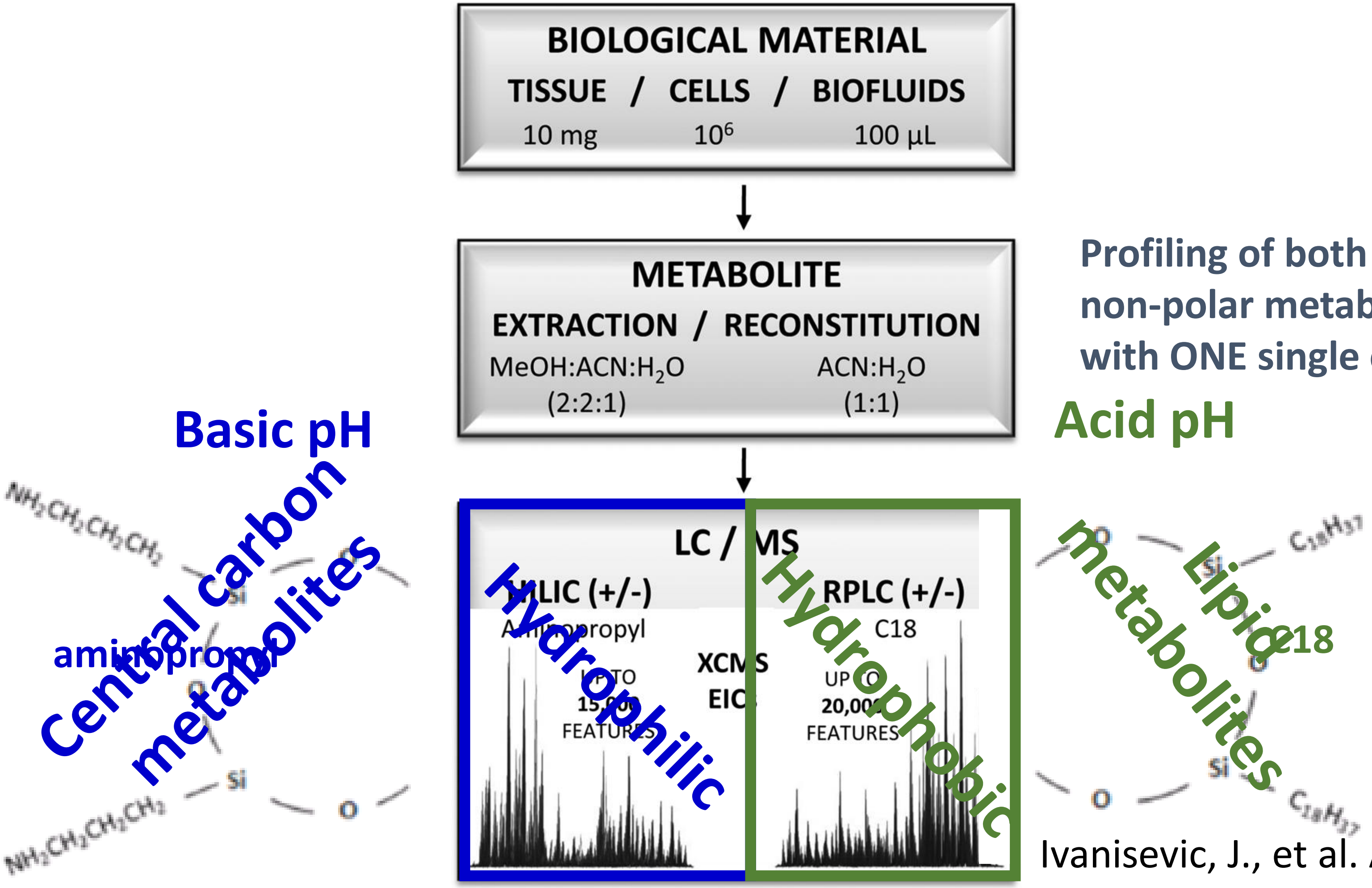
Acid pH

aminopropyl

C18

Ivanisevic, J., et al. Anal. Chem,

Fundamental Metabolomics



BIOLOGICAL MATERIAL
TISSUE / CELLS / BIOFLUIDS
10 mg 10^6 100 µL

METABOLITE
EXTRACTION / RECONSTITUTION
$MeOH:ACN:H_2O$ (2:2:1) $ACN:H_2O$ (1:1)

LC / MS
HILIC (+/-) RPLC (+/-)
Aminopropyl C18
XCMS
EIC
UP TO 15,000 FEATURES
UP TO 20,000 FEATURES

Profiling of both polar and non-polar metabolites with ONE single extraction

Basic pH

Acid pH

Central carbon metabolites

aminopropyl

Hydrophilic

Hydrophobic

Lipid metabolites

C18

Ivanisevic, J., et al. Anal. Chem,

Fundamental Metabolomics

LC-MS based Untargeted Metabolomics

- Choice of MS
- LC-MS combination
- **Analytical sequence**
- Data processing
- Metabolite identification

LC-MS Analytical sequence

Blank (50/50 ACN/H2O)

Standard mixture (mixture of metabolite standards)

QC sample (pooled sample solution)

Sample 1 - 10

QC sample

Sample 11 - 20

⋮

Examples of Large-Scale Metabolomic Profiling

•Sreekumar et al. (Nature 2009) used a 16-min gradient to examine 262 prostate-related samples (tissue, plasma, urine) to look for markers of prostate cancer aggressiveness

•Wang et al. (Nature Medicine 2011) used a 30-min run to analyze 1500 plasma samples for predictors of diabetes

•Wang et al (Nature 2011) used 14.5-min run to analyze 2000 plasma samples for cardiovascular disease.

Fundamental Metabolomics

LC-MS based Untargeted Metabolomics

- Choice of MS
- LC-MS combination
- Analytical sequence
- **Data processing**
- Metabolite identification

Fundamental Metabolomics

XCMS Online

Web-based metabolomics data processing platform

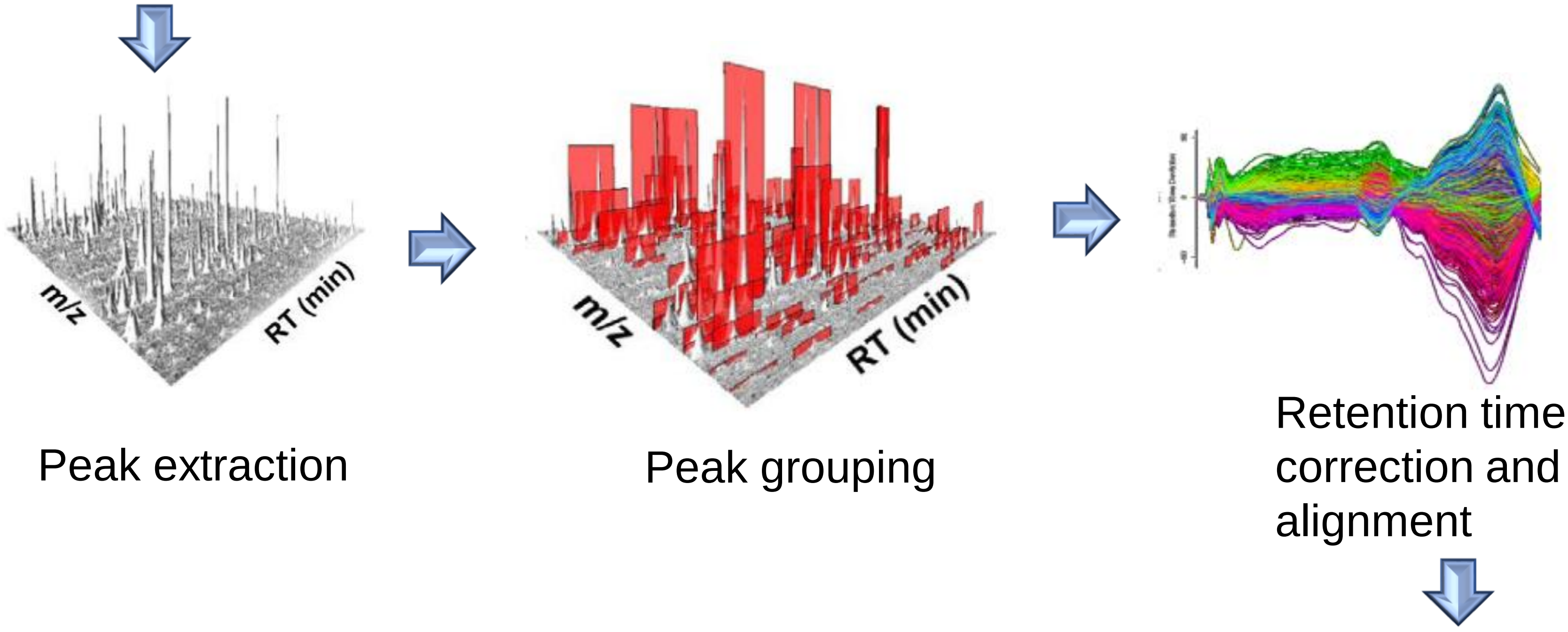
Fundamental Metabolomics

XCMS Online

Web-based metabolomics data processing platform

- Comprehensive data processing algorithms
- Interactive tools for rapid data exploration
- Over 15,000 registered users (May, 2018)
- Freely available at http://xcmsonline.scripps.edu/

Fundamental Metabolomics



Raw MS data

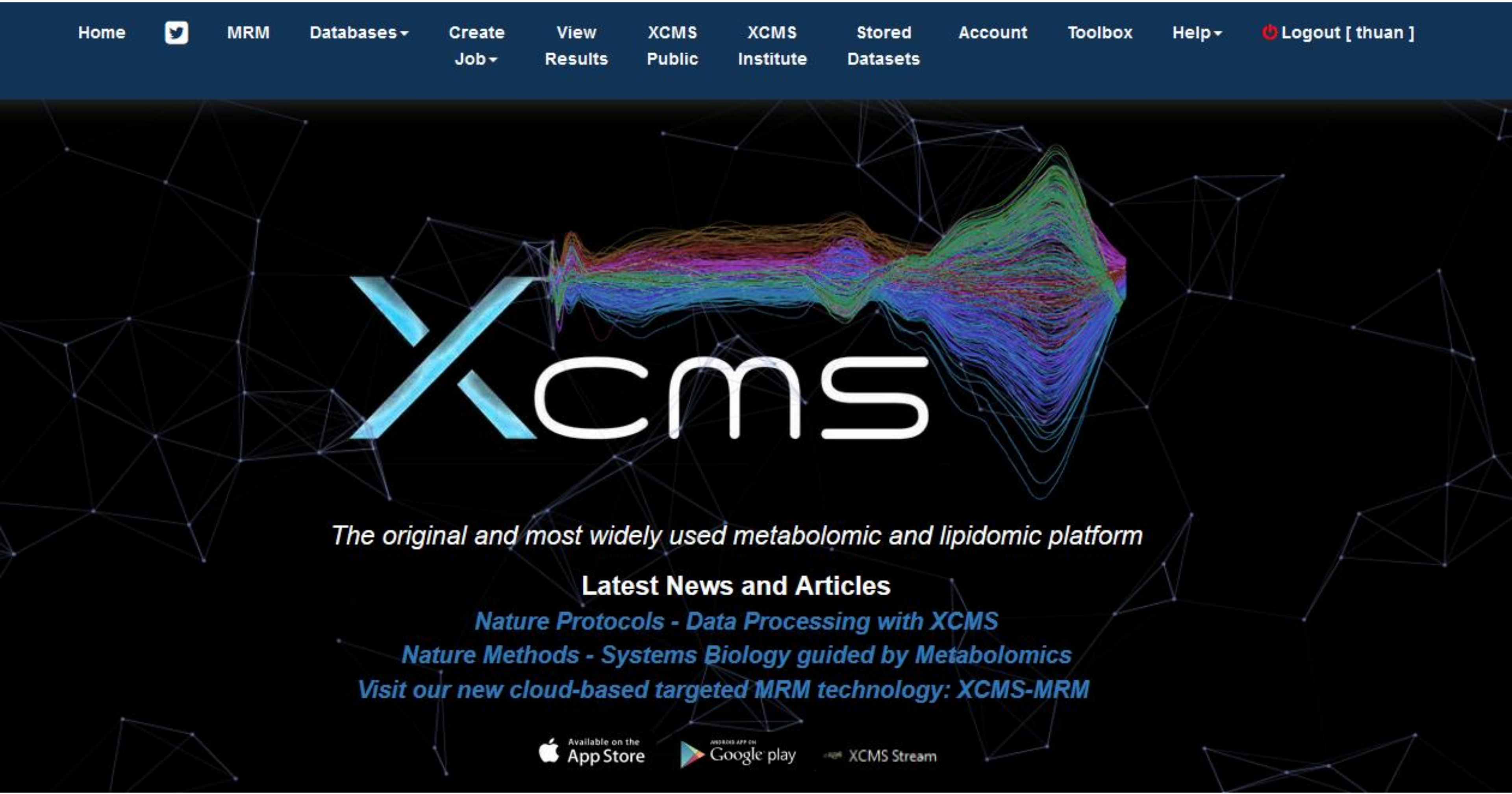
Peak extraction

Peak grouping

Retention time correction and alignment

Metabolic feature table

Fundamental Metabolomics



The original and most widely used metabolomic and lipidomic platform

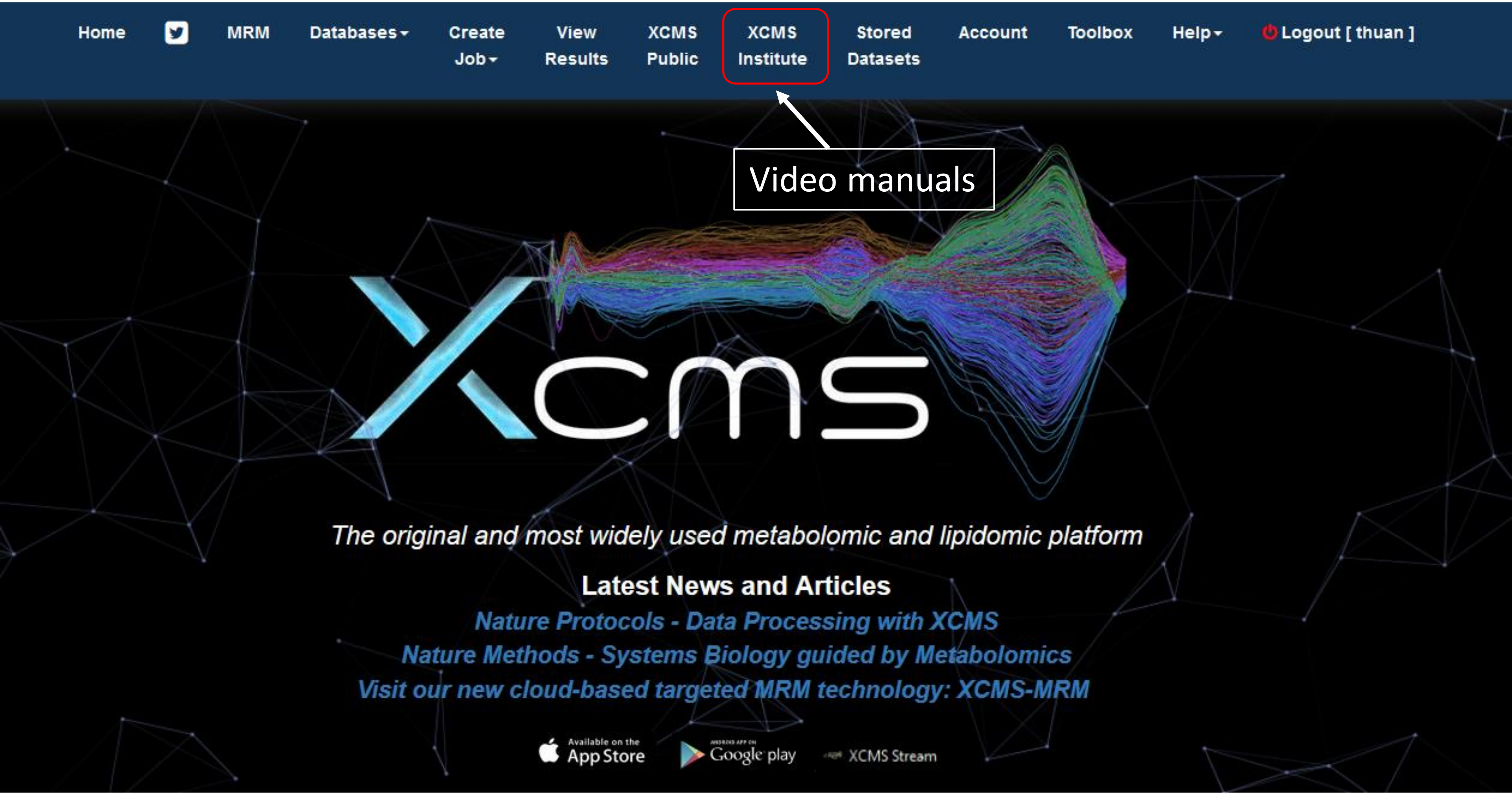
Latest News and Articles

Nature Protocols - Data Processing with XCMS

Nature Methods - Systems Biology guided by Metabolomics

Visit our new cloud-based targeted MRM technology: XCMS-MRM

Fundamental Metabolomics



Video manuals

The original and most widely used metabolomic and lipidomic platform

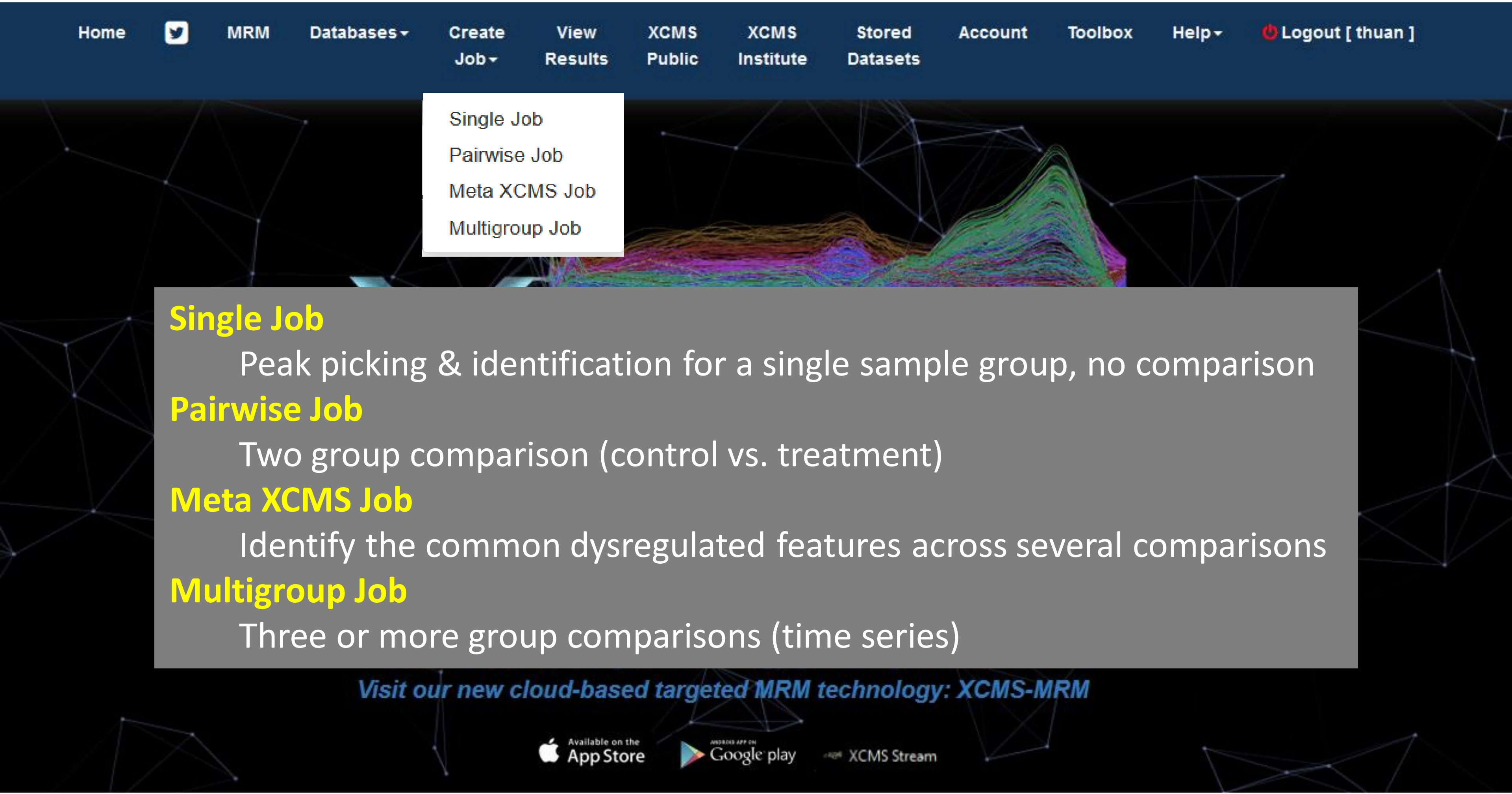
Latest News and Articles

Nature Protocols - Data Processing with XCMS

Nature Methods - Systems Biology guided by Metabolomics

Visit our new cloud-based targeted MRM technology: XCMS-MRM

Fundamental Metabolomics



Single Job

Peak picking & identification for a single sample group, no comparison

Pairwise Job

Two group comparison (control vs. treatment)

Meta XCMS Job

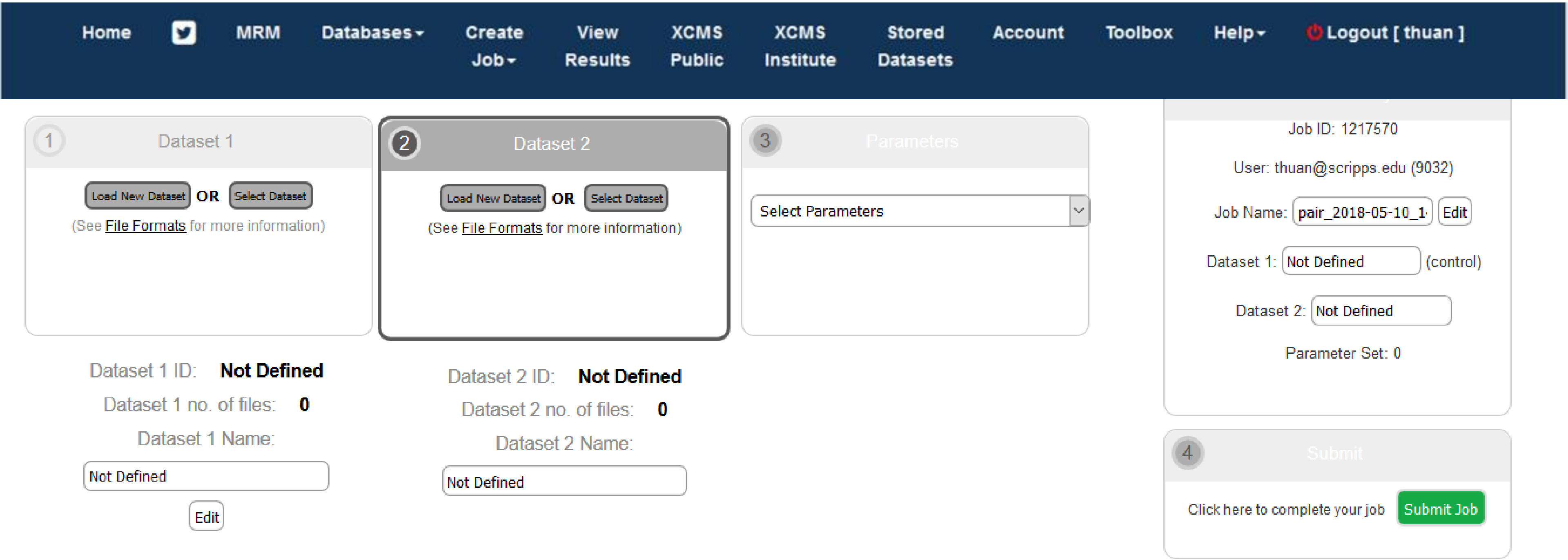
Identify the common dysregulated features across several comparisons

Multigroup Job

Three or more group comparisons (time series)

Visit our new cloud-based targeted MRM technology: XCMS-MRM

Fundamental Metabolomics



(See File Formats for more information)

(See File Formats for more information)

Dataset 1 ID: **Not Defined**

Dataset 1 no. of files: **0**

Dataset 1 Name: Not Defined

Dataset 2 ID: **Not Defined**

Dataset 2 no. of files: **0**

Dataset 2 Name: Not Defined

Job ID: 1217570

User: thuan@scripps.edu (9032)

Job Name: pair_2018-05-10_1

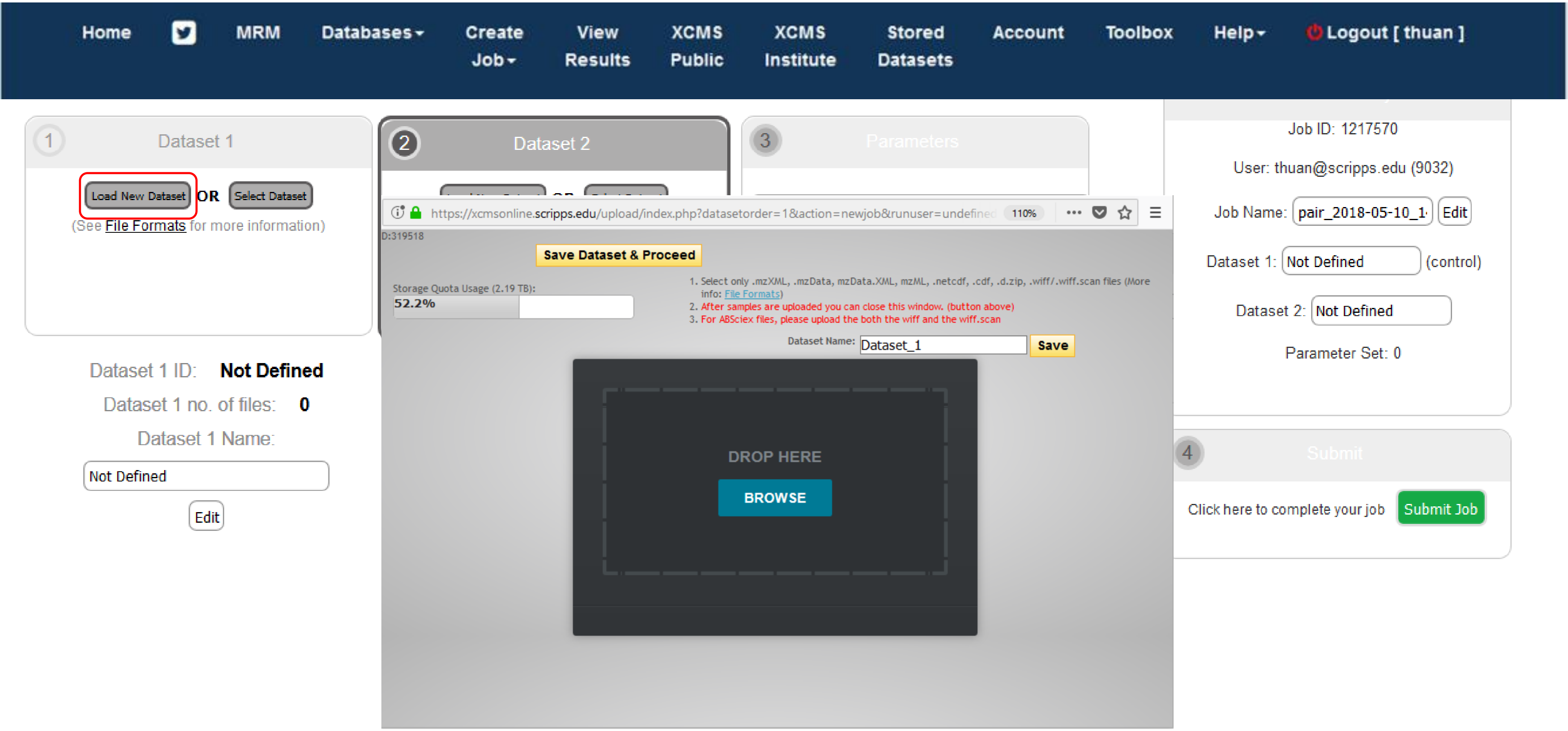
Dataset 1: Not Defined (control)

Dataset 2: Not Defined

Parameter Set: 0

Click here to complete your job

Fundamental Metabolomics



(See File Formats for more information)

Dataset 1 ID: **Not Defined**

Dataset 1 no. of files: **0**

Dataset 1 Name: Not Defined

https://xcmsonline.scripps.edu/upload/index.php?datasetorder=1&action=newjob&runuser=undefined 110%

D:319518

Storage Quota Usage (2.19 TB): 52.2%

1. Select only .mzXML, .mzData, mzData.XML, mzML, .netcdf, .cdf, .d.zip, .wiff/.wiff.scan files (More info: File Formats)
2. After samples are uploaded you can close this window. (button above)
3. For ABSciex files, please upload the both the wiff and the wiff.scan

Dataset Name: Dataset_1

Job ID: 1217570

User: thuan@scripps.edu (9032)

Job Name: pair_2018-05-10_1

Dataset 1: Not Defined (control)

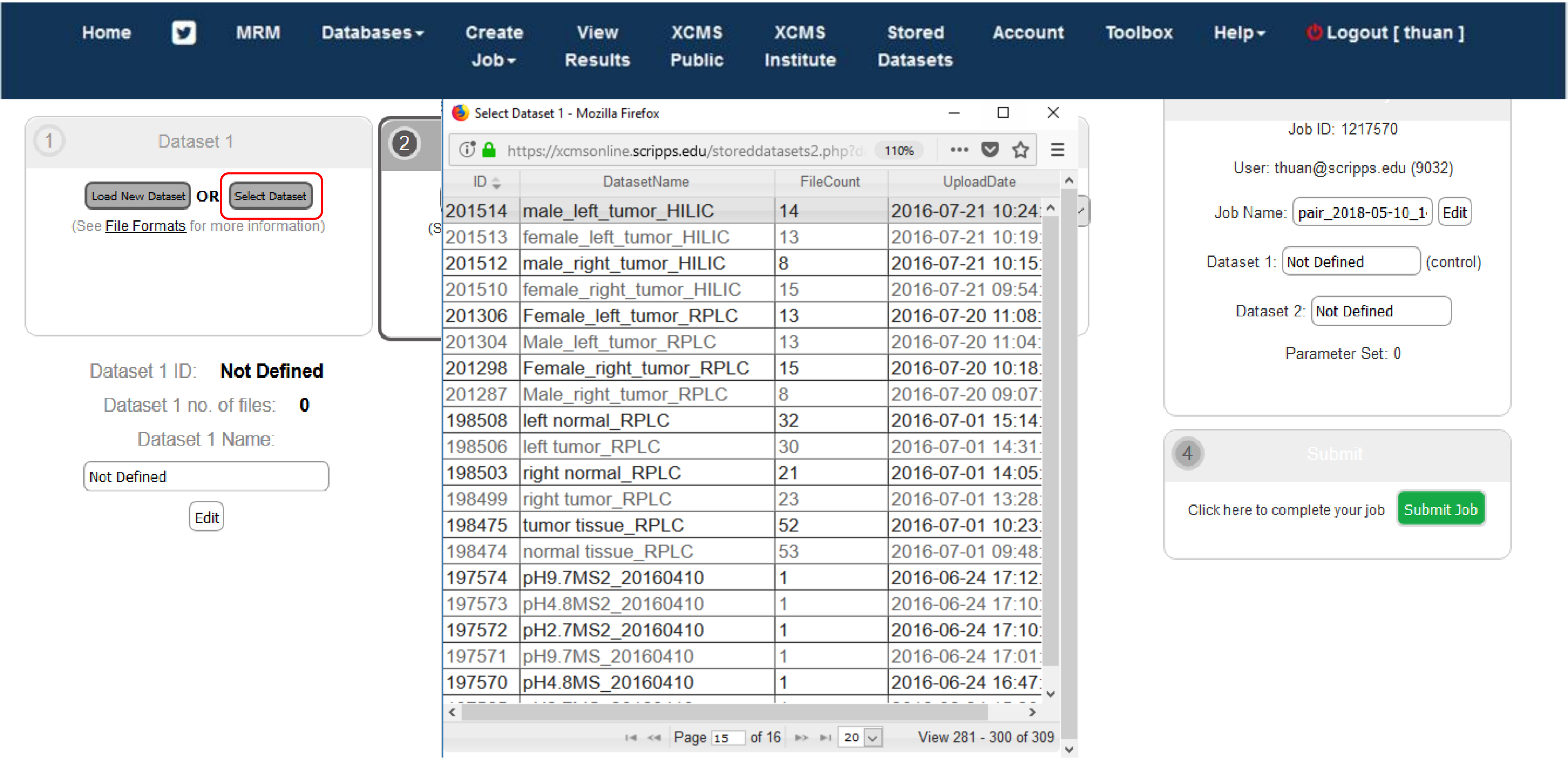
Dataset 2: Not Defined

Parameter Set: 0

Click here to complete your job

Directly upload new data files

Fundamental Metabolomics



(See File Formats for more information)

Dataset 1 ID: **Not Defined**

Dataset 1 no. of files: **0**

Dataset 1 Name: Not Defined

Select Dataset 1 - Mozilla Firefox

https://xcmsonline.scripps.edu/storeddatasets2.php?d 110%

ID	DatasetName	FileCount	UploadDate
201514	male_left_tumor_HILIC	14	2016-07-21 10:24:
201513	female_left_tumor_HILIC	13	2016-07-21 10:19:
201512	male_right_tumor_HILIC	8	2016-07-21 10:15:
201510	female_right_tumor_HILIC	15	2016-07-21 09:54:
201306	Female_left_tumor_RPLC	13	2016-07-20 11:08:
201304	Male_left_tumor_RPLC	13	2016-07-20 11:04:
201298	Female_right_tumor_RPLC	15	2016-07-20 10:18:
201287	Male_right_tumor_RPLC	8	2016-07-20 09:07:
198508	left normal_RPLC	32	2016-07-01 15:14:
198506	left tumor_RPLC	30	2016-07-01 14:31:
198503	right normal_RPLC	21	2016-07-01 14:05:
198499	right tumor_RPLC	23	2016-07-01 13:28:
198475	tumor tissue_RPLC	52	2016-07-01 10:23:
198474	normal tissue_RPLC	53	2016-07-01 09:48:
197574	pH9.7MS2_20160410	1	2016-06-24 17:12:
197573	pH4.8MS2_20160410	1	2016-06-24 17:10:
197572	pH2.7MS2_20160410	1	2016-06-24 17:10:
197571	pH9.7MS_20160410	1	2016-06-24 17:01:
197570	pH4.8MS_20160410	1	2016-06-24 16:47:

Page 15 of 16 20 View 281 - 300 of 309

Job ID: 1217570

User: thuan@scripps.edu (9032)

Job Name: pair_2018-05-10_1

Dataset 1: Not Defined (control)

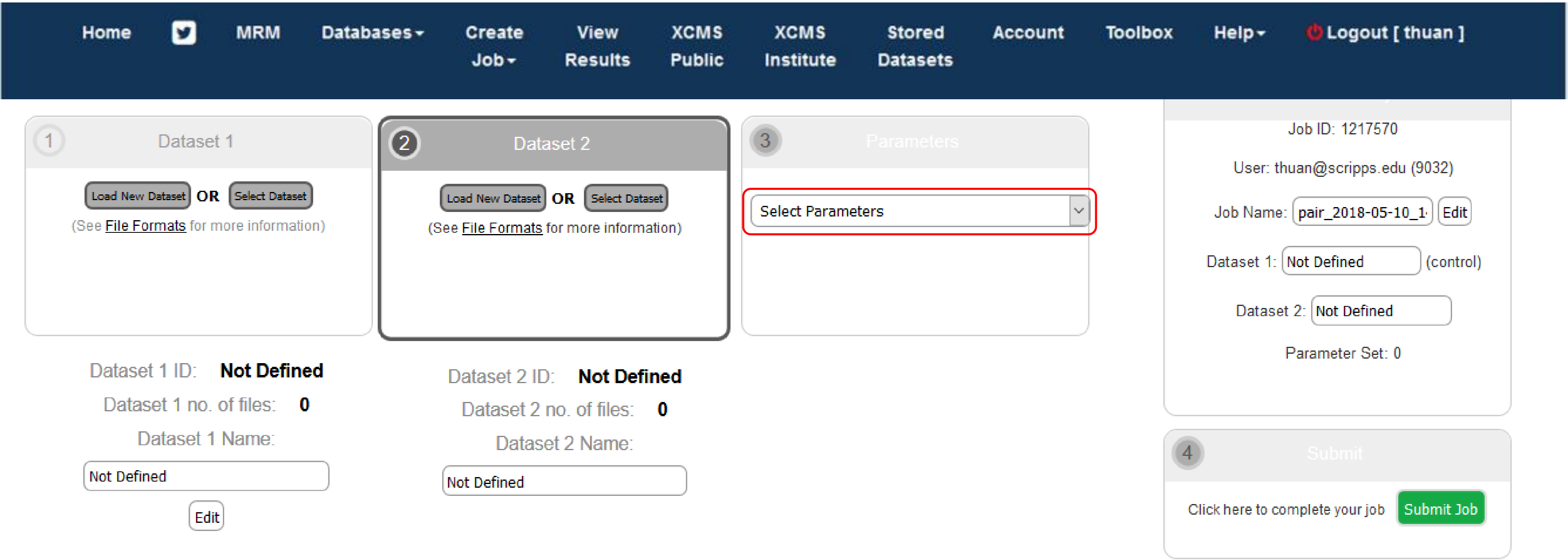
Dataset 2: Not Defined

Parameter Set: 0

Click here to complete your job

Select from a list of existing stored datasets

Fundamental Metabolomics



(See File Formats for more information)

(See File Formats for more information)

Dataset 1 ID: **Not Defined**

Dataset 1 no. of files: **0**

Dataset 1 Name: Not Defined

Dataset 2 ID: **Not Defined**

Dataset 2 no. of files: **0**

Dataset 2 Name: Not Defined

Job ID: 1217570

User: thuan@scripps.edu (9032)

Job Name: pair_2018-05-10_1

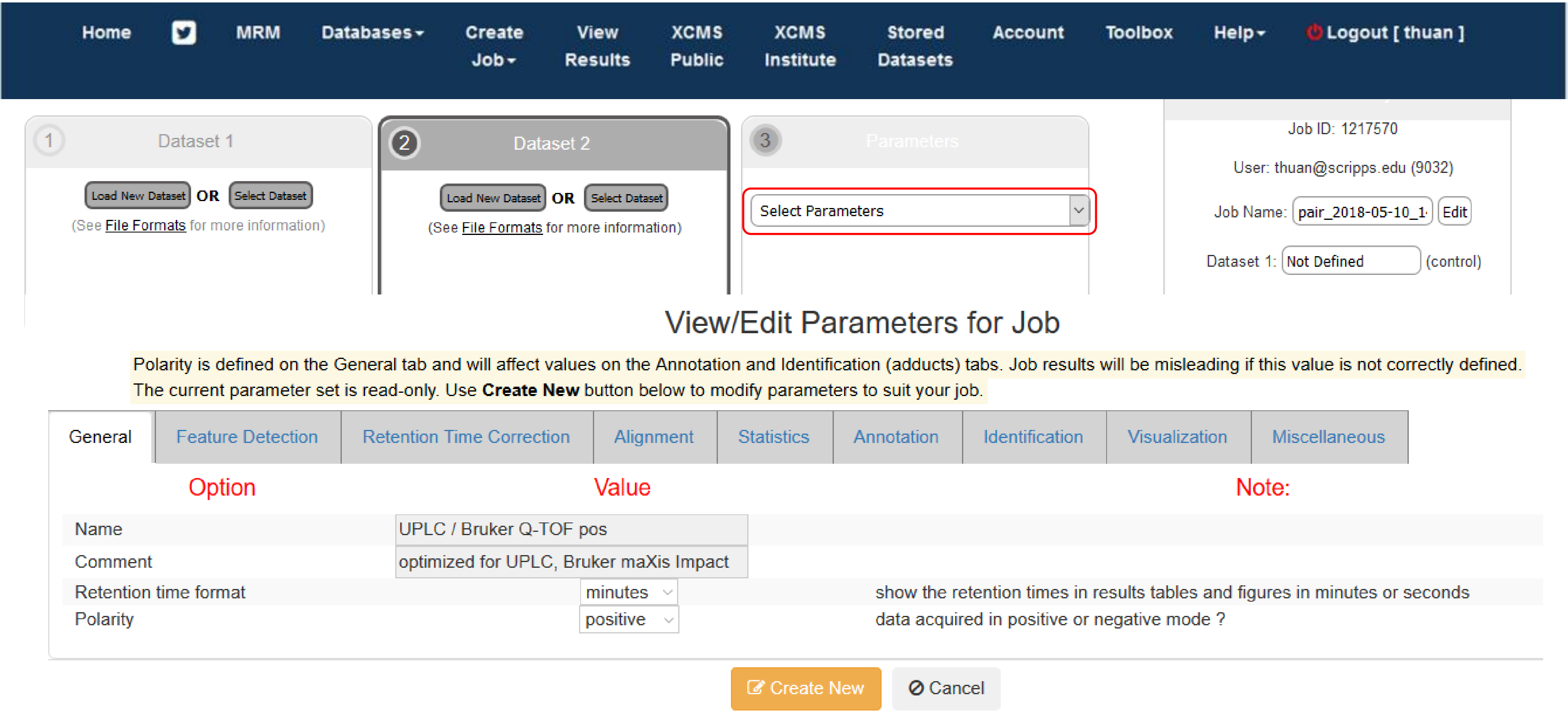
Dataset 1: Not Defined (control)

Dataset 2: Not Defined

Parameter Set: 0

Click here to complete your job

Fundamental Metabolomics



Job ID: 1217570

User: thuan@scripps.edu (9032)

Job Name: pair_2018-05-10_1

Dataset 1: Not Defined (control)

View/Edit Parameters for Job

Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.
The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

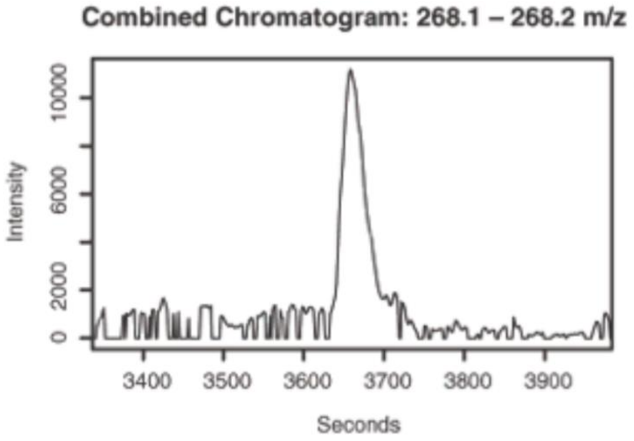
Option	Value	Note:
Name	UPLC / Bruker Q-TOF pos	
Comment	optimized for UPLC, Bruker maXis Impact	
Retention time format	minutes	show the retention times in results tables and figures in minutes or seconds
Polarity	positive	data acquired in positive or negative mode ?

Fundamental Metabolomics

Simplified mechanism of peak extraction

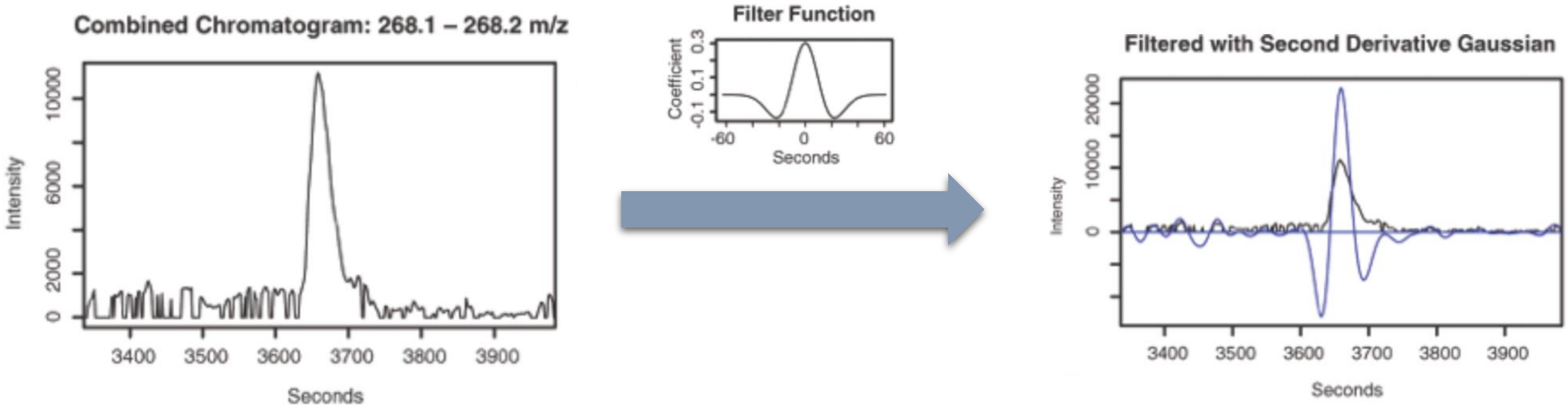
Fundamental Metabolomics

Raw LC-MS data



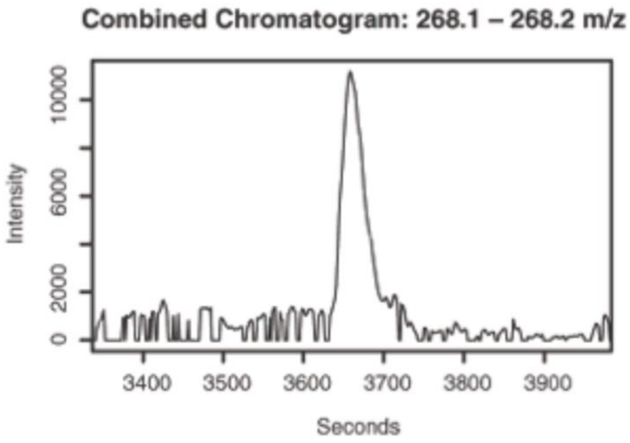
Fundamental Metabolomics

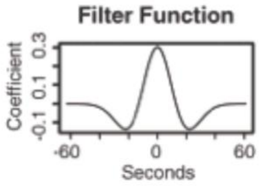
Raw LC-MS data

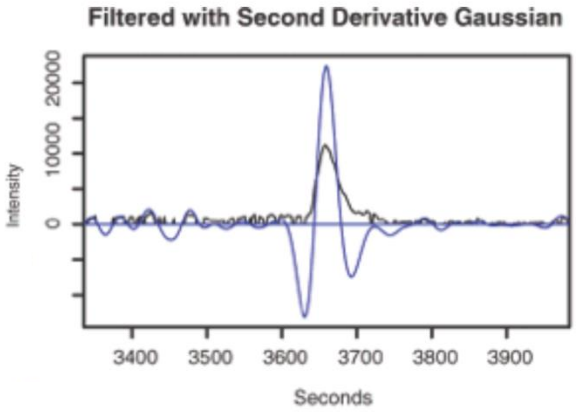


Fundamental Metabolomics

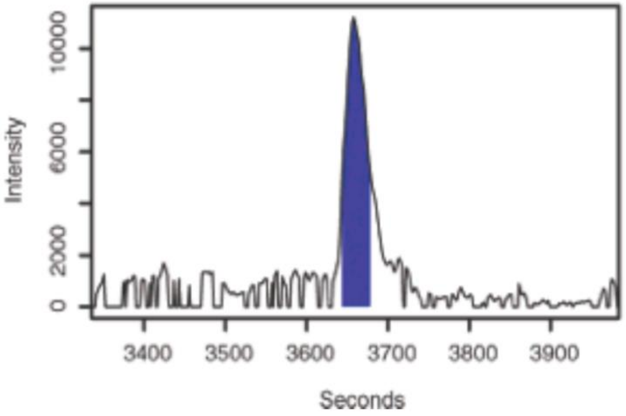
Raw LC-MS data





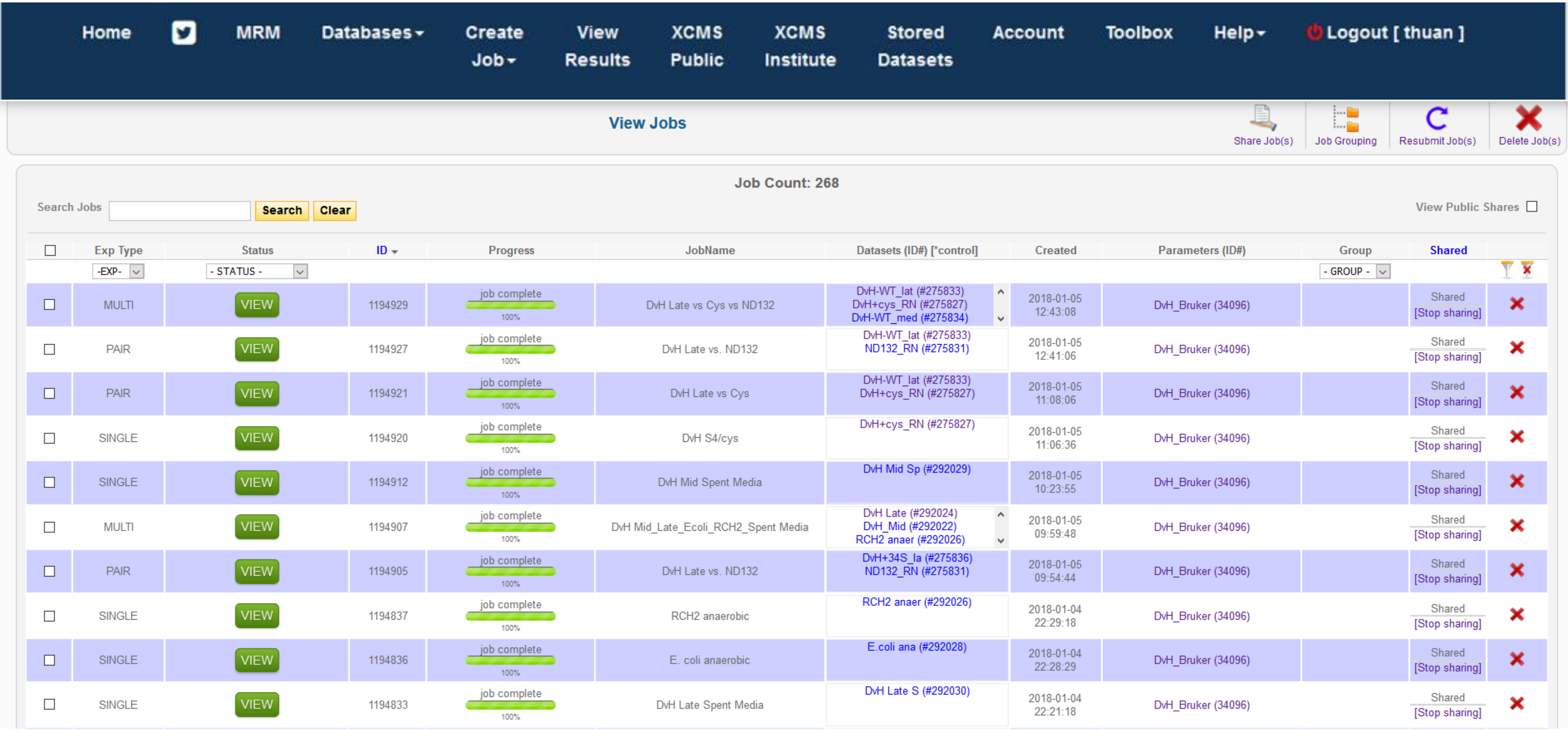


LC-MS peak



Mechanism of peak extraction

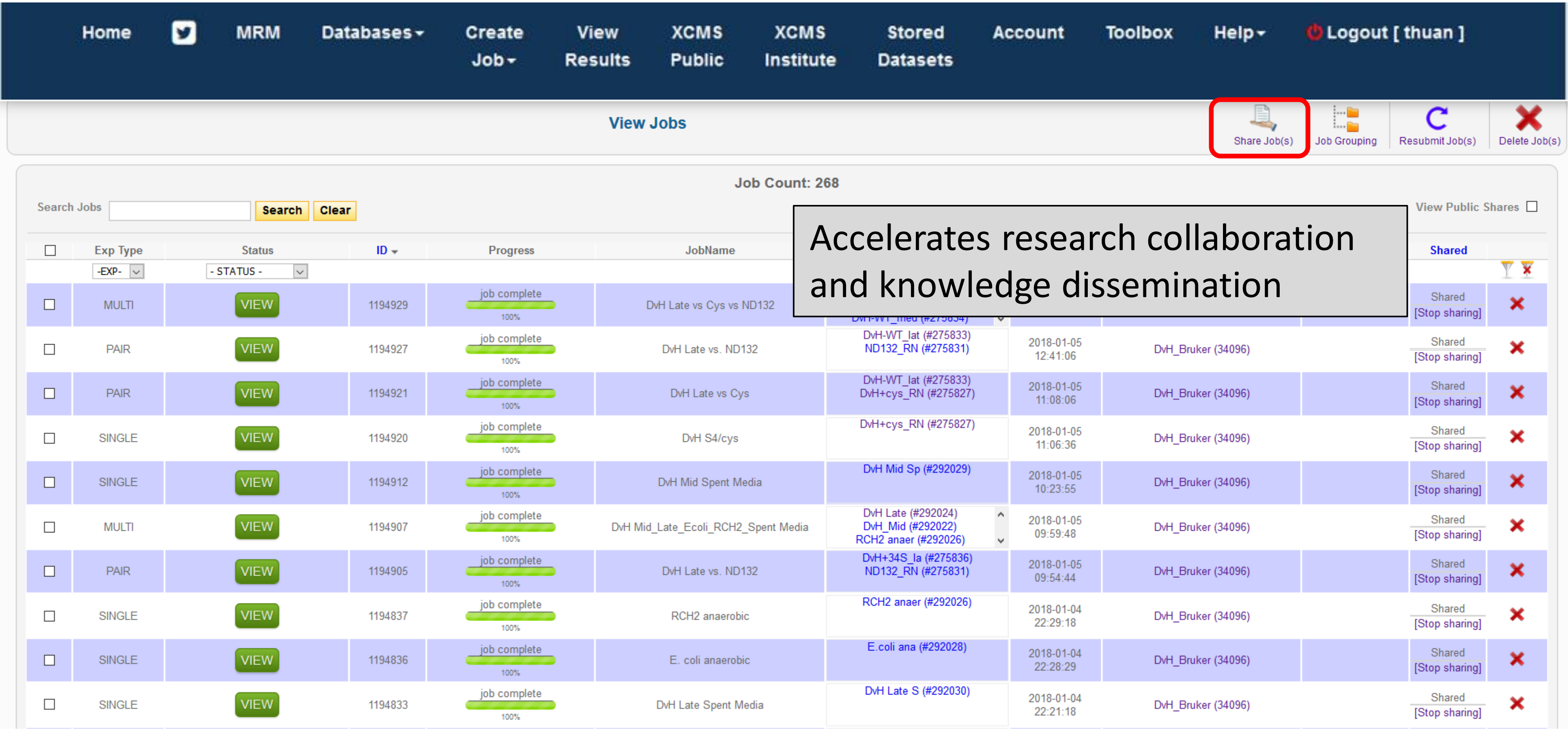
Fundamental Metabolomics



Job Count: 268

☐	Exp Type	Status	ID	Progress	JobName	Datasets (ID#) [*control]	Created	Parameters (ID#)	Group	Shared	
	-EXP-	- STATUS -							- GROUP -		
☐	MULTI	VIEW	1194929	job complete 100%	DvH Late vs Cys vs ND132	DvH-WT_lat (#275833) DvH+cys_RN (#275827) DvH-WT_med (#275834)	2018-01-05 12:43:08	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	PAIR	VIEW	1194927	job complete 100%	DvH Late vs. ND132	DvH-WT_lat (#275833) ND132_RN (#275831)	2018-01-05 12:41:06	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	PAIR	VIEW	1194921	job complete 100%	DvH Late vs Cys	DvH-WT_lat (#275833) DvH+cys_RN (#275827)	2018-01-05 11:08:06	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	SINGLE	VIEW	1194920	job complete 100%	DvH S4/cys	DvH+cys_RN (#275827)	2018-01-05 11:06:36	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	SINGLE	VIEW	1194912	job complete 100%	DvH Mid Spent Media	DvH Mid Sp (#292029)	2018-01-05 10:23:55	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	MULTI	VIEW	1194907	job complete 100%	DvH Mid_Late_Ecoli_RCH2_Spent Media	DvH Late (#292024) DvH_Mid (#292022) RCH2 anaer (#292026)	2018-01-05 09:59:48	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	PAIR	VIEW	1194905	job complete 100%	DvH Late vs. ND132	DvH+34S_la (#275836) ND132_RN (#275831)	2018-01-05 09:54:44	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	SINGLE	VIEW	1194837	job complete 100%	RCH2 anaerobic	RCH2 anaer (#292026)	2018-01-04 22:29:18	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	SINGLE	VIEW	1194836	job complete 100%	E. coli anaerobic	E.coli ana (#292028)	2018-01-04 22:28:29	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	SINGLE	VIEW	1194833	job complete 100%	DvH Late Spent Media	DvH Late S (#292030)	2018-01-04 22:21:18	DvH_Bruker (34096)		Shared [Stop sharing]	✖

Fundamental Metabolomics

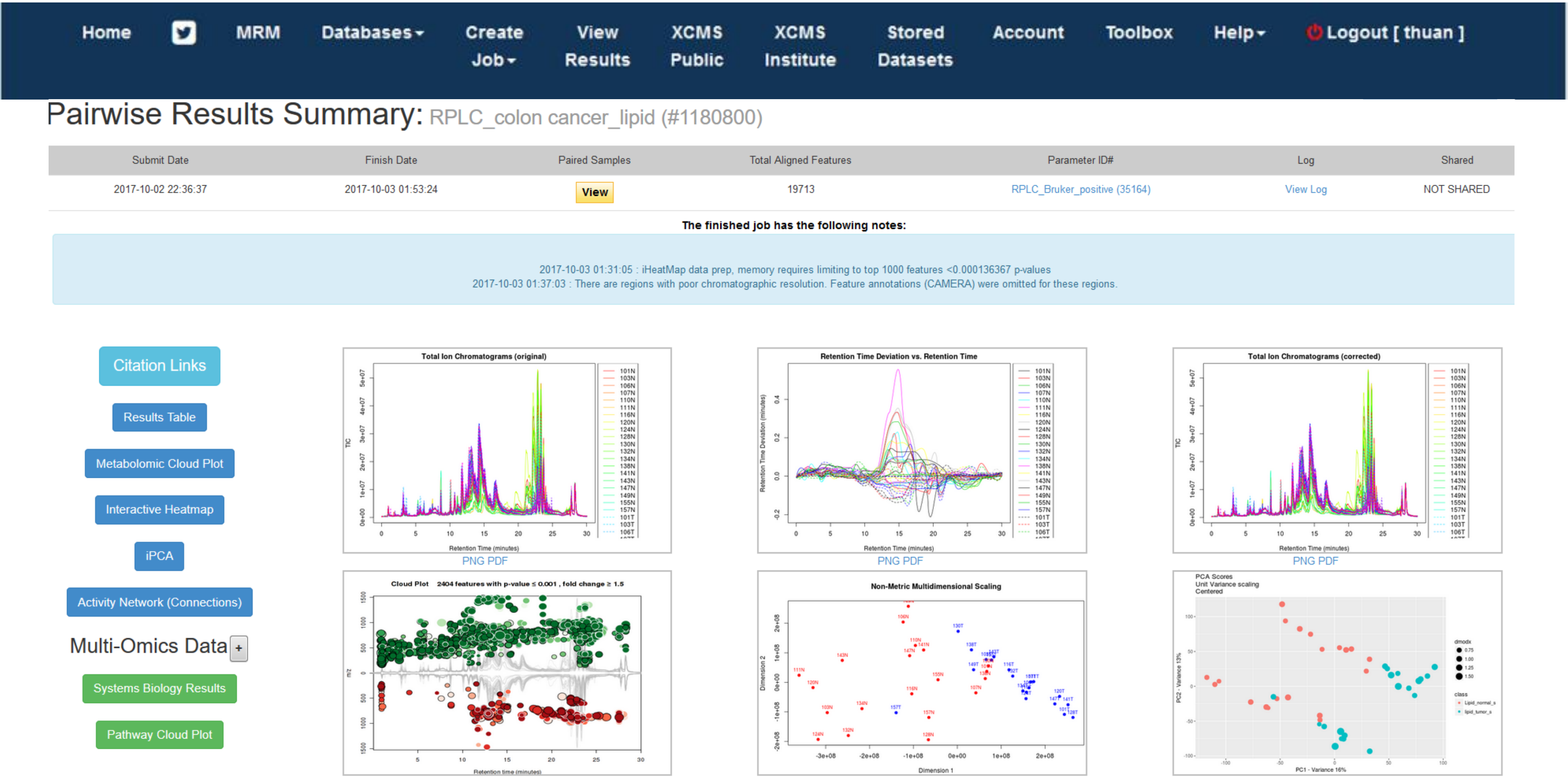


Job Count: 268

☐	Exp Type	Status	ID	Progress	JobName					Shared	
☐	MULTI	VIEW	1194929	job complete 100%	DvH Late vs Cys vs ND132	DvH-WT_med (#275834)				Shared [Stop sharing]	✖
☐	PAIR	VIEW	1194927	job complete 100%	DvH Late vs. ND132	DvH-WT_lat (#275833) ND132_RN (#275831)	2018-01-05 12:41:06	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	PAIR	VIEW	1194921	job complete 100%	DvH Late vs Cys	DvH-WT_lat (#275833) DvH+cys_RN (#275827)	2018-01-05 11:08:06	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	SINGLE	VIEW	1194920	job complete 100%	DvH S4/cys	DvH+cys_RN (#275827)	2018-01-05 11:06:36	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	SINGLE	VIEW	1194912	job complete 100%	DvH Mid Spent Media	DvH Mid Sp (#292029)	2018-01-05 10:23:55	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	MULTI	VIEW	1194907	job complete 100%	DvH Mid_Late_Ecoli_RCH2_Spent Media	DvH Late (#292024) DvH_Mid (#292022) RCH2 anaer (#292026)	2018-01-05 09:59:48	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	PAIR	VIEW	1194905	job complete 100%	DvH Late vs. ND132	DvH+34S_la (#275836) ND132_RN (#275831)	2018-01-05 09:54:44	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	SINGLE	VIEW	1194837	job complete 100%	RCH2 anaerobic	RCH2 anaer (#292026)	2018-01-04 22:29:18	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	SINGLE	VIEW	1194836	job complete 100%	E. coli anaerobic	E.coli ana (#292028)	2018-01-04 22:28:29	DvH_Bruker (34096)		Shared [Stop sharing]	✖
☐	SINGLE	VIEW	1194833	job complete 100%	DvH Late Spent Media	DvH Late S (#292030)	2018-01-04 22:21:18	DvH_Bruker (34096)		Shared [Stop sharing]	✖

Accelerates research collaboration and knowledge dissemination

Fundamental Metabolomics



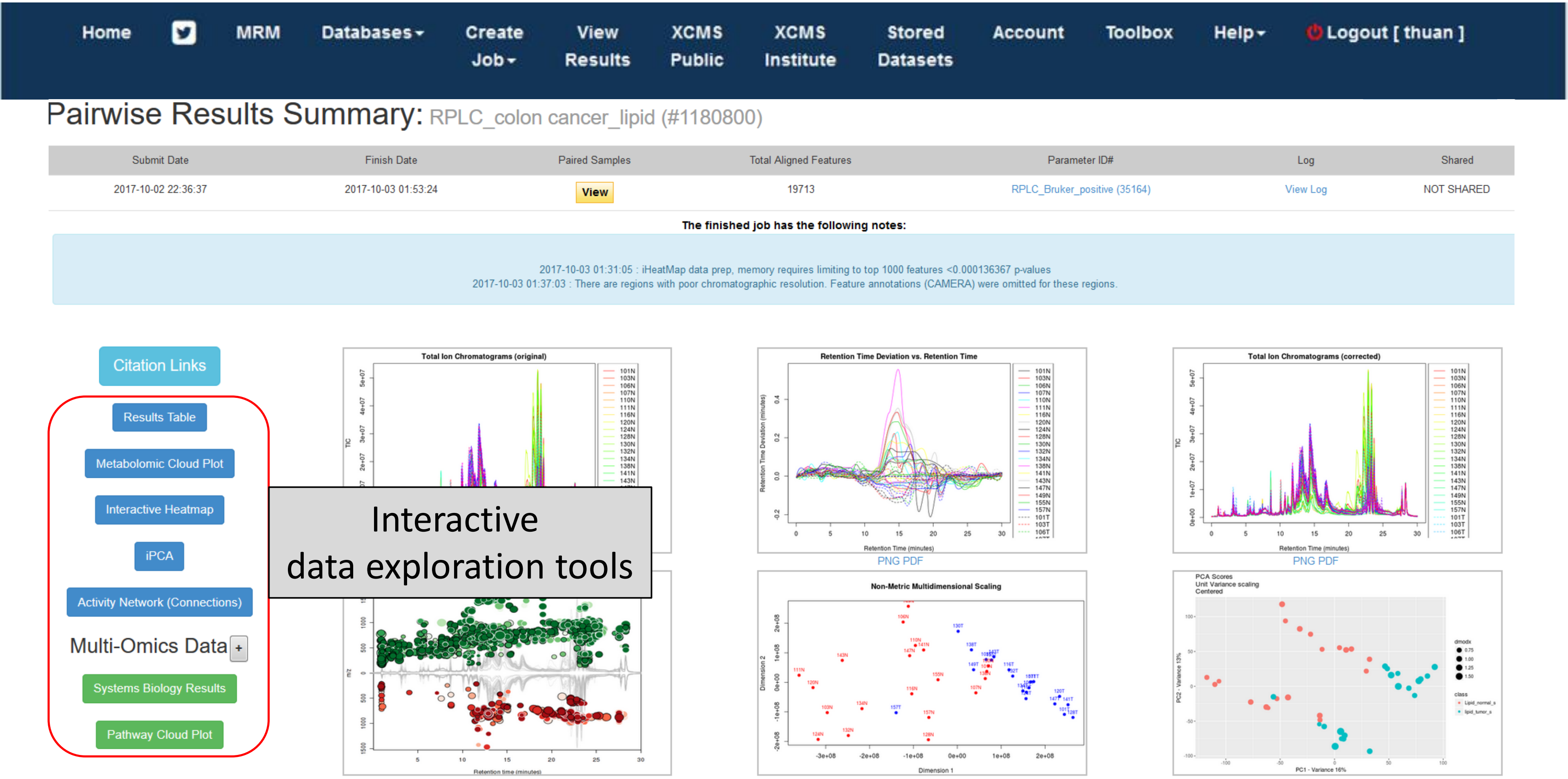
Pairwise Results Summary: RPLC_colon cancer_lipid (#1180800)

Submit Date	Finish Date	Paired Samples	Total Aligned Features	Parameter ID#	Log	Shared
2017-10-02 22:36:37	2017-10-03 01:53:24	View	19713	RPLC_Bruker_positive (35164)	View Log	NOT SHARED

The finished job has the following notes:

2017-10-03 01:31:05 : iHeatMap data prep, memory requires limiting to top 1000 features <0.000136367 p-values
2017-10-03 01:37:03 : There are regions with poor chromatographic resolution. Feature annotations (CAMERA) were omitted for these regions.

Fundamental Metabolomics



Pairwise Results Summary: RPLC_colon cancer_lipid (#1180800)

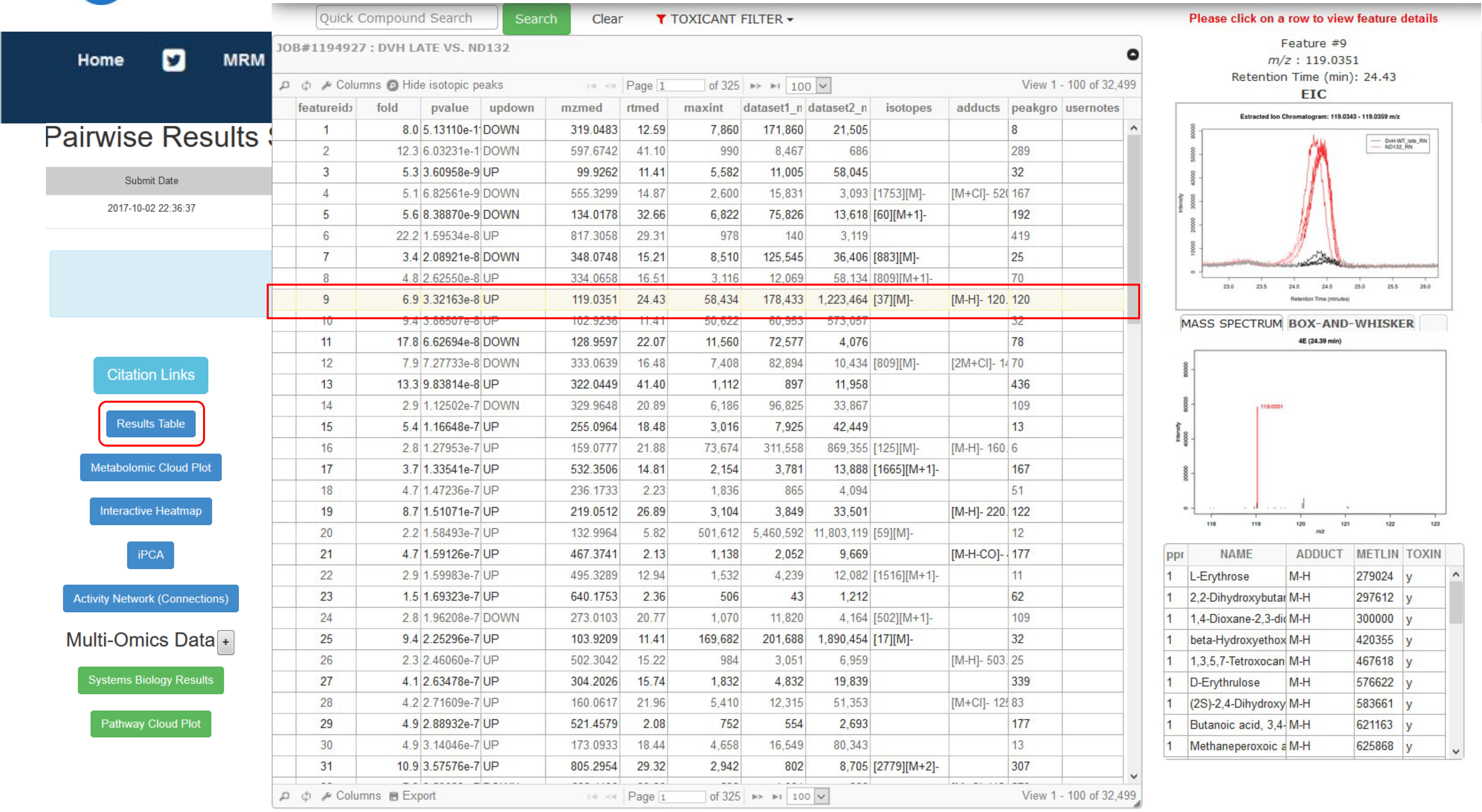
Submit Date	Finish Date	Paired Samples	Total Aligned Features	Parameter ID#	Log	Shared
2017-10-02 22:36:37	2017-10-03 01:53:24	View	19713	RPLC_Bruker_positive (35164)	View Log	NOT SHARED

The finished job has the following notes:

2017-10-03 01:31:05 : iHeatMap data prep, memory requires limiting to top 1000 features <0.000136367 p-values
2017-10-03 01:37:03 : There are regions with poor chromatographic resolution. Feature annotations (CAMERA) were omitted for these regions.

Interactive
data exploration tools

Fundamental Metabolomics



Please click on a row to view feature details

JOB#1194927 : DVH LATE VS. ND132

featureid›	fold	pvalue	updown	mzmed	rtmed	maxint	dataset1_n	dataset2_n	isotopes	adducts	peakgro	usernotes
1	8.0	5.13110e-1	DOWN	319.0483	12.59	7,860	171,860	21,505			8	
2	12.3	6.03231e-1	DOWN	597.6742	41.10	990	8,467	686			289	
3	5.3	3.60958e-9	UP	99.9262	11.41	5,582	11,005	58,045			32	
4	5.1	6.82561e-9	DOWN	555.3299	14.87	2,600	15,831	3,093	[1753][M]-	[M+Cl]- 52	167	
5	5.6	8.38870e-9	DOWN	134.0178	32.66	6,822	75,826	13,618	[60][M+1]-		192	
6	22.2	1.59534e-8	UP	817.3058	29.31	978	140	3,119			419	
7	3.4	2.08921e-8	DOWN	348.0748	15.21	8,510	125,545	36,406	[883][M]-		25	
8	4.8	2.62550e-8	UP	334.0658	16.51	3,116	12,069	58,134	[809][M+1]-		70	
9	6.9	3.32163e-8	UP	119.0351	24.43	58,434	178,433	1,223,464	[37][M]-	[M-H]- 120.	120	
10	9.4	3.86507e-8	UP	102.9236	11.41	50,622	60,953	573,057			32	
11	17.8	6.62694e-8	DOWN	128.9597	22.07	11,560	72,577	4,076			78	
12	7.9	7.27733e-8	DOWN	333.0639	16.48	7,408	82,894	10,434	[809][M]-	[2M+Cl]- 14	70	
13	13.3	9.83814e-8	UP	322.0449	41.40	1,112	897	11,958			436	
14	2.9	1.12502e-7	DOWN	329.9648	20.89	6,186	96,825	33,867			109	
15	5.4	1.16648e-7	UP	255.0964	18.48	3,016	7,925	42,449			13	
16	2.8	1.27953e-7	UP	159.0777	21.88	73,674	311,558	869,355	[125][M]-	[M-H]- 160.	6	
17	3.7	1.33541e-7	UP	532.3506	14.81	2,154	3,781	13,888	[1665][M+1]-		167	
18	4.7	1.47236e-7	UP	236.1733	2.23	1,836	865	4,094			51	
19	8.7	1.51071e-7	UP	219.0512	26.89	3,104	3,849	33,501		[M-H]- 220.	122	
20	2.2	1.58493e-7	UP	132.9964	5.82	501,612	5,460,592	11,803,119	[59][M]-		12	
21	4.7	1.59126e-7	UP	467.3741	2.13	1,138	2,052	9,669		[M-H-CO]-	177	
22	2.9	1.59983e-7	UP	495.3289	12.94	1,532	4,239	12,082	[1516][M+1]-		11	
23	1.5	1.69323e-7	UP	640.1753	2.36	506	43	1,212			62	
24	2.8	1.96208e-7	DOWN	273.0103	20.77	1,070	11,820	4,164	[502][M+1]-		109	
25	9.4	2.25296e-7	UP	103.9209	11.41	169,682	201,688	1,890,454	[17][M]-		32	
26	2.3	2.46060e-7	UP	502.3042	15.22	984	3,051	6,959		[M-H]- 503.	25	
27	4.1	2.63478e-7	UP	304.2026	15.74	1,832	4,832	19,839			339	
28	4.2	2.71609e-7	UP	160.0617	21.96	5,410	12,315	51,353		[M+Cl]- 12	83	
29	4.9	2.88932e-7	UP	521.4579	2.08	752	554	2,693			177	
30	4.9	3.14046e-7	UP	173.0933	18.44	4,658	16,549	80,343			13	
31	10.9	3.57576e-7	UP	805.2954	29.32	2,942	802	8,705	[2779][M+2]-		307	

Feature #9
m/z : 119.0351
Retention Time (min): 24.43
EIC

ppr	NAME	ADDUCT	METLIN	TOXIN
1	L-Erythrose	M-H	279024	y
1	2,2-Dihydroxybuta	M-H	297612	y
1	1,4-Dioxane-2,3-di	M-H	300000	y
1	beta-Hydroxyethox	M-H	420355	y
1	1,3,5,7-Tetroxocan	M-H	467618	y
1	D-Erythrulose	M-H	576622	y
1	(2S)-2,4-Dihydroxy	M-H	583661	y
1	Butanoic acid, 3,4-	M-H	621163	y
1	Methaneperoxoic a	M-H	625868	y

Fundamental Metabolomics

LC-MS based Untargeted Metabolomics

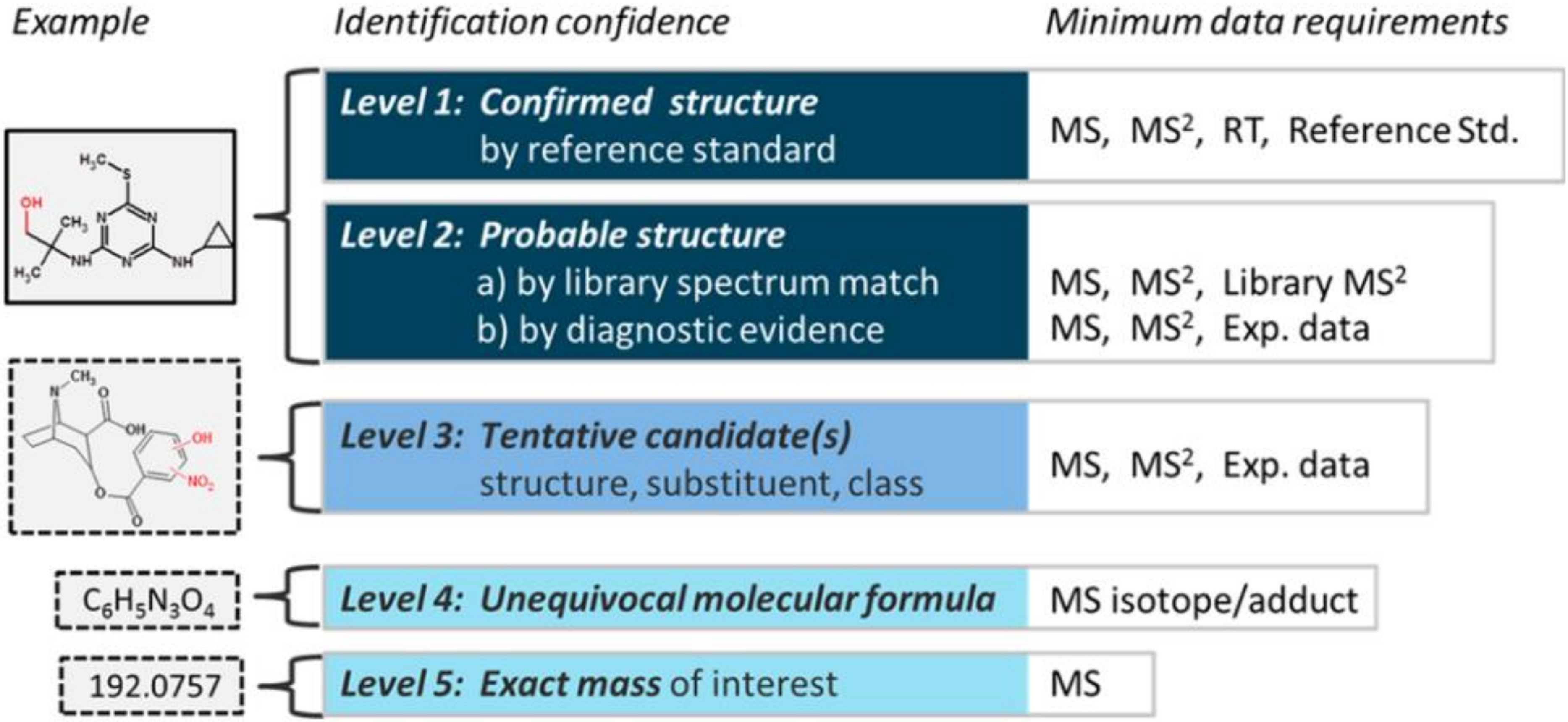
- Choice of MS
- LC-MS combination
- Analytical sequence
- Data processing
- **Metabolite identification**

Fundamental Metabolomics

Metabolite identification

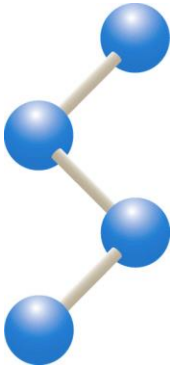
- Accurate mass
- Retention time
- MS/MS spectra

Fundamental Metabolomics



Example	Identification confidence	Minimum data requirements
	Level 1: Confirmed structure by reference standard	MS, MS^2, RT, Reference Std.
	Level 2: Probable structure a) by library spectrum match b) by diagnostic evidence	MS, MS^2, Library MS^2 MS, MS^2, Exp. data
	Level 3: Tentative candidate(s) structure, substituent, class	MS, MS^2, Exp. data
$C_6H_5N_3O_4$	***Level 4: Unequivocal molecular formula***	MS isotope/adduct
192.0757	***Level 5: Exact mass*** of interest	MS

Adapted from Schymanski et. al., *Environ. Sci. Technol.* 2014, 48, 2097-2098



Fundamental Metabolomics

Databases

Fundamental Metabolomics

Databases

METLIN

KEGG

hmdb
The Human Metabolome Database

LIPID MAPS™

NIST
National Institute of
Standards and Technology

MassBank

Fundamental Metabolomics

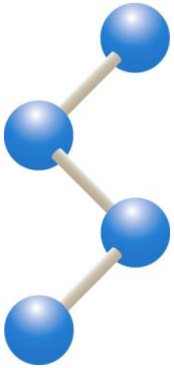
Databases

- The largest MS/MS collection of data at multiple collision energies and in positive and negative ionization modes.
- It now includes 961,829 molecules ranging from lipids, steroids, plant & bacteria metabolites, small peptides, carbohydrates, exogenous drugs/metabolites, central carbon metabolites and toxicants.
- Over 14,000 metabolites have been individually analyzed and another 200,000 has in silico MS/MS data.

Fundamental Metabolomics

Databases

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from metabolomic information.



Fundamental Metabolomics

Databases

- A freely available electronic database containing detailed information about small molecule metabolites found in the human body.
- 42,003 metabolite entries including both water-soluble and lipid soluble metabolites as well as metabolites that would be regarded as either abundant or relatively rare.

hmdb

The Human Metabolome Database

Databases

LIPID Metabolites And Pathways Strategy (LIPID MAPS) is a multi-institutional effort to identify and quantify lipd species in mammalian cells, as well as to identify the changes in response to perturbation.

- Detailed lipid pathways
- Analytical tools and experimental protocols
- 500 mass spectrometric standards
- Experimental data

Fundamental Metabolomics

Databases

The NIST Chemistry WebBook provides chemical and physical property data for chemical species. The data provided are from collections maintained by the NIST Standard Reference Data Program and outside contributors.

NIST
National Institute of
Standards and Technology

Fundamental Metabolomics

Databases

- A public repository of mass spectra of small chemical compounds
- It contains spectra from EI-MS, FAB-MS, ESI-MS^2

MassBank

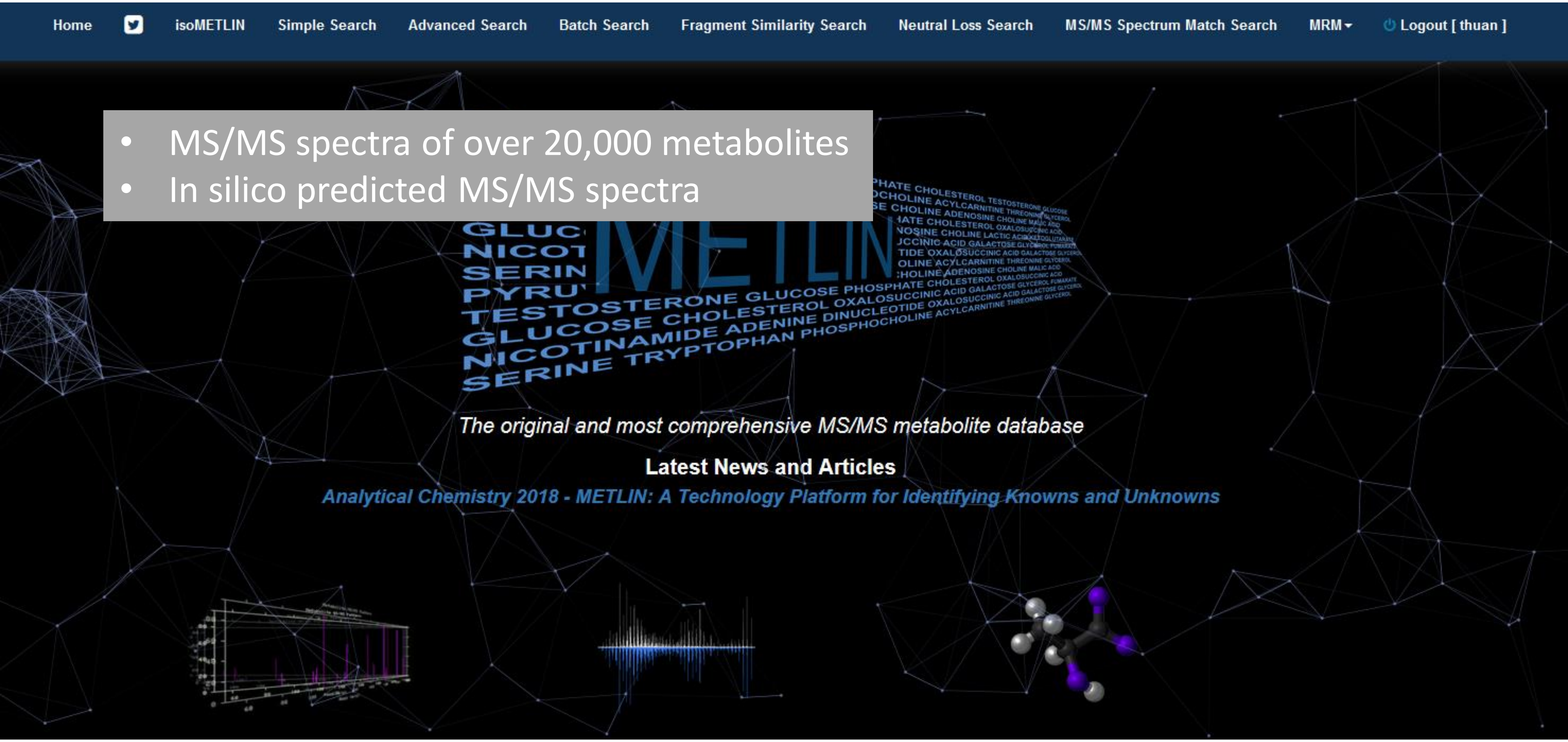
Fundamental Metabolomics

The original and most comprehensive MS/MS metabolite database

Latest News and Articles

Analytical Chemistry 2018 - METLIN: A Technology Platform for Identifying Knowns and Unknowns

Fundamental Metabolomics



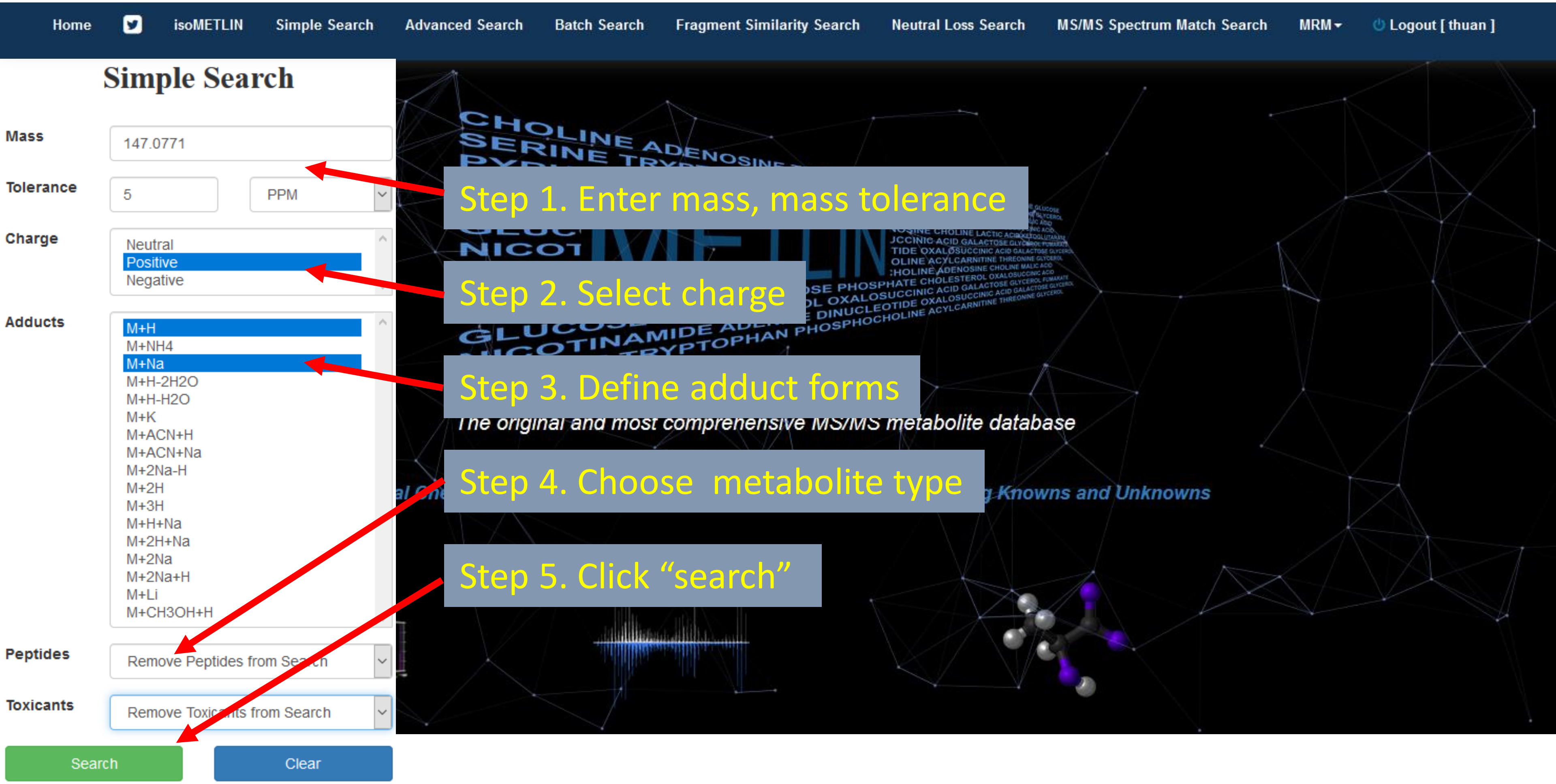
- MS/MS spectra of over 20,000 metabolites
- In silico predicted MS/MS spectra

The original and most comprehensive MS/MS metabolite database

Latest News and Articles

Analytical Chemistry 2018 - METLIN: A Technology Platform for Identifying Knowns and Unknowns

Fundamental Metabolomics



Simple Search

Mass 147.0771

Tolerance 5 PPM

Charge
- Neutral
- Positive (selected)
- Negative

Adducts
- M+H (selected)
- M+NH4
- M+Na (selected)
- 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 Search

Toxicants Remove Toxicants from Search

Step 1. Enter mass, mass tolerance

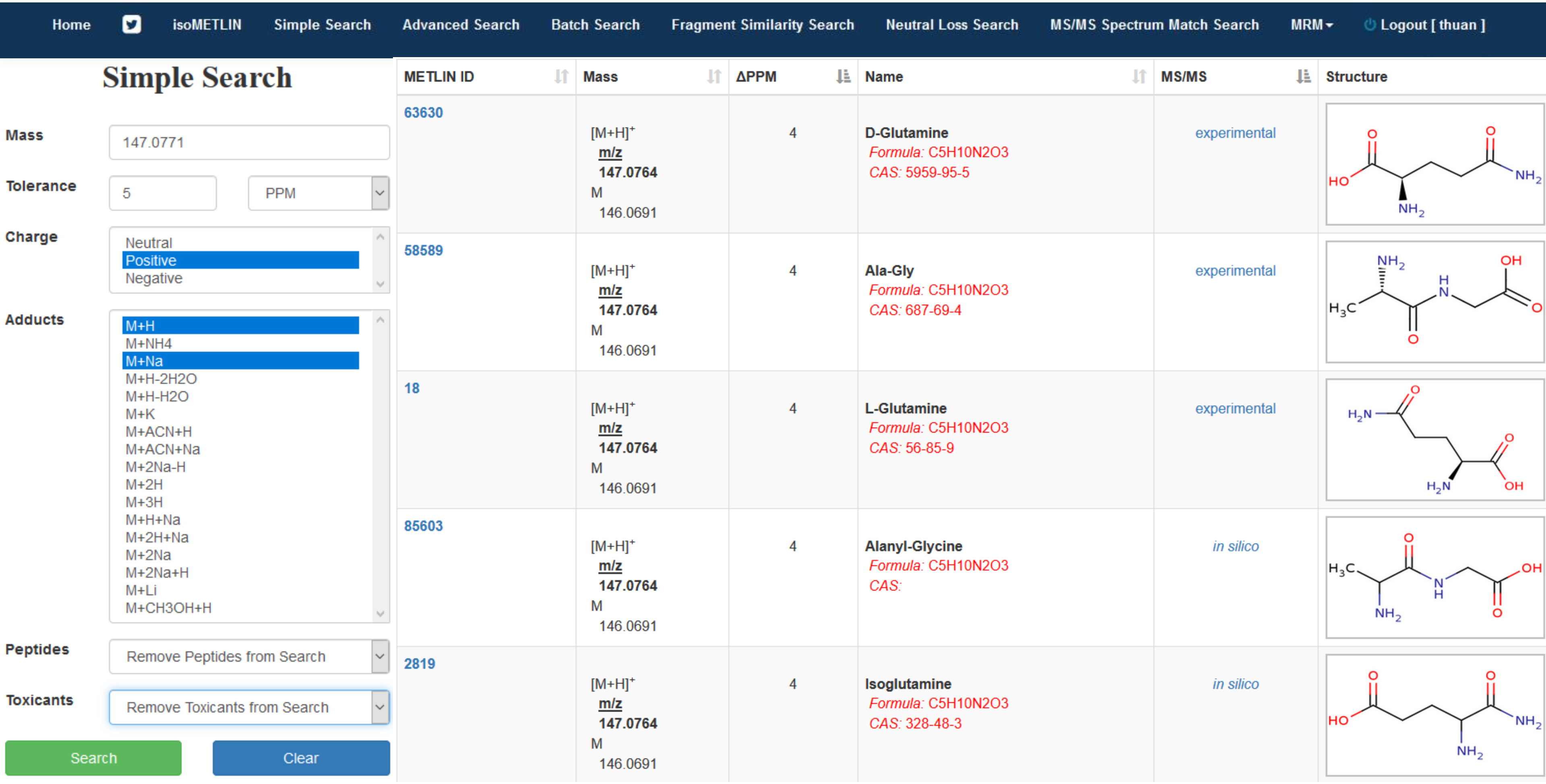
Step 2. Select charge

Step 3. Define adduct forms

Step 4. Choose metabolite type

Step 5. Click “search”

Fundamental Metabolomics



Simple Search

Mass 147.0771

Tolerance 5 PPM

Charge Neutral [✓ Positive] Negative

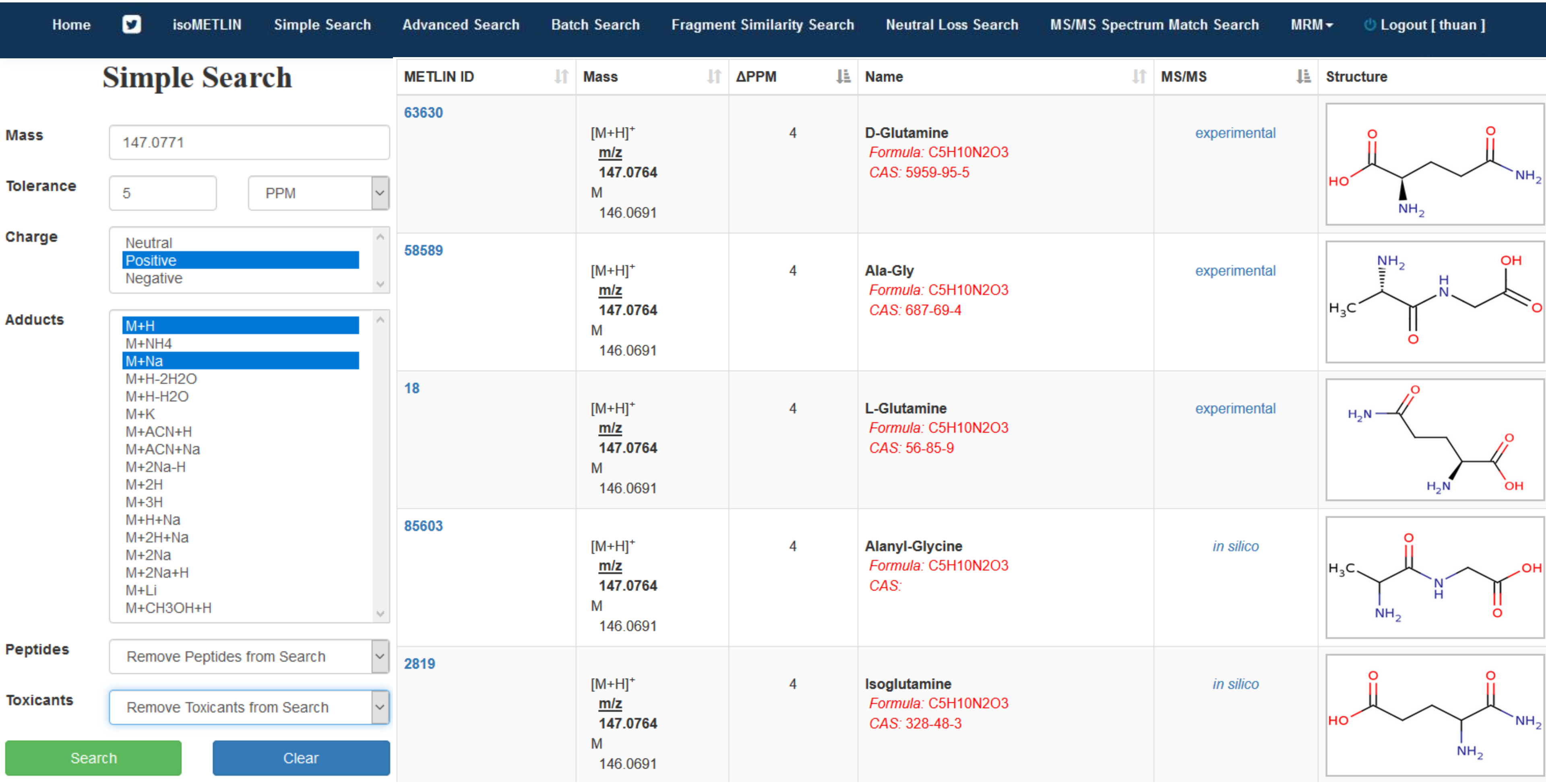
Adducts
- [✓ M+H]
- M+NH4
- [✓ M+Na]
- M+H-2H2O
- M+H-H2O
- M+K
- M+ACN+H
- M+ACN+Na
- M+2Na-H
- M+2H
- M+3H
- M+H+Na
- M+2H+Na
- M+2Na
- M+2Na+H
- M+Li
- M+CH3OH+H

Peptides Remove Peptides from Search

Toxicants Remove Toxicants from Search

METLIN ID	Mass	ΔPPM	Name	MS/MS	Structure
63630	[M+H]$^+$ **m/z** **147.0764** M 146.0691	4	**D-Glutamine** *Formula:* C5H10N2O3 *CAS:* 5959-95-5	experimental	
58589	[M+H]$^+$ **m/z** **147.0764** M 146.0691	4	**Ala-Gly** *Formula:* C5H10N2O3 *CAS:* 687-69-4	experimental	
18	[M+H]$^+$ **m/z** **147.0764** M 146.0691	4	**L-Glutamine** *Formula:* C5H10N2O3 *CAS:* 56-85-9	experimental	
85603	[M+H]$^+$ **m/z** **147.0764** M 146.0691	4	**Alanyl-Glycine** *Formula:* C5H10N2O3 *CAS:*	*in silico*	
2819	[M+H]$^+$ **m/z** **147.0764** M 146.0691	4	**Isoglutamine** *Formula:* C5H10N2O3 *CAS:* 328-48-3	*in silico*	

Fundamental Metabolomics



Simple Search

Mass 147.0771

Tolerance 5 PPM

Charge
- Neutral
- Positive
- Negative

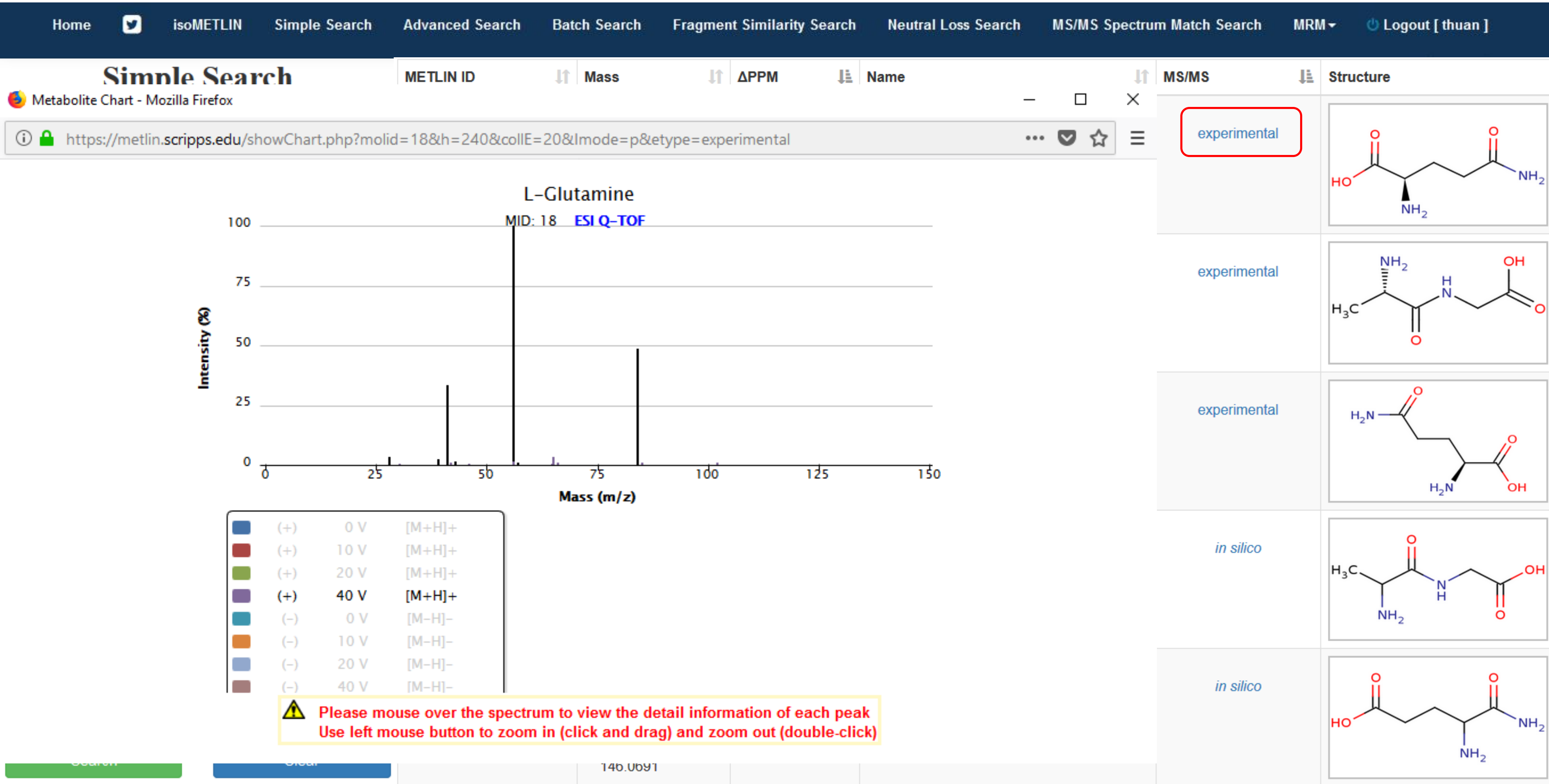
Adducts
- M+H
- M+NH4
- M+Na
- M+H-2H2O
- M+H-H2O
- M+K
- M+ACN+H
- M+ACN+Na
- M+2Na-H
- M+2H
- M+3H
- M+H+Na
- M+2H+Na
- M+2Na
- M+2Na+H
- M+Li
- M+CH3OH+H

Peptides Remove Peptides from Search

Toxicants Remove Toxicants from Search

METLIN ID	Mass	ΔPPM	Name	MS/MS	Structure
63630	$[M+H]^+$ m/z 147.0764 M 146.0691	4	**D-Glutamine** *Formula:* C5H10N2O3 *CAS:* 5959-95-5	experimental	
58589	$[M+H]^+$ m/z 147.0764 M 146.0691	4	**Ala-Gly** *Formula:* C5H10N2O3 *CAS:* 687-69-4	experimental	
18	$[M+H]^+$ m/z 147.0764 M 146.0691	4	**L-Glutamine** *Formula:* C5H10N2O3 *CAS:* 56-85-9	experimental	
85603	$[M+H]^+$ m/z 147.0764 M 146.0691	4	**Alanyl-Glycine** *Formula:* C5H10N2O3 *CAS:*	*in silico*	
2819	$[M+H]^+$ m/z 147.0764 M 146.0691	4	**Isoglutamine** *Formula:* C5H10N2O3 *CAS:* 328-48-3	*in silico*	

Fundamental Metabolomics



Simple Search

METLIN ID	Mass	ΔPPM	Name	MS/MS	Structure
				experimental	
				experimental	
				experimental	
				in silico	
				in silico	

146.0691

https://metlin.scripps.edu/showChart.php?molid=18&h=240&collE=20&lmode=p&etype=experimental

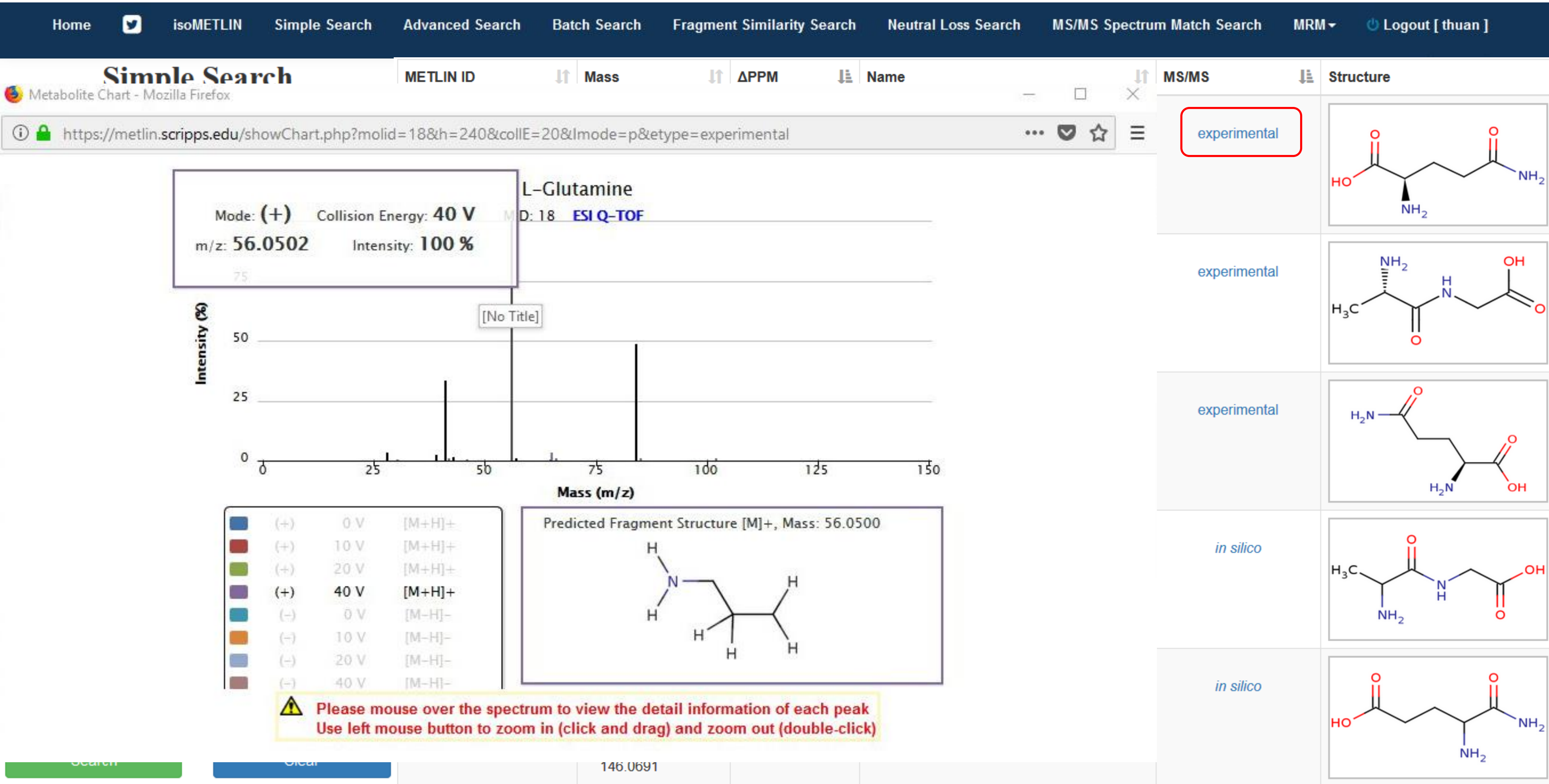
L-Glutamine

MID: 18 ESI Q-TOF

🟦	(+)	0 V	[M+H]+
🟥	(+)	10 V	[M+H]+
🟩	(+)	20 V	[M+H]+
🟪	(+)	40 V	[M+H]+
🟦	(-)	0 V	[M-H]-
🟧	(-)	10 V	[M-H]-
🟦	(-)	20 V	[M-H]-
🟫	(-)	40 V	[M-H]-

⚠ Please mouse over the spectrum to view the detail information of each peak
Use left mouse button to zoom in (click and drag) and zoom out (double-click)

Fundamental Metabolomics



METLIN ID	Mass	ΔPPM	Name	MS/MS	Structure
				experimental	
				experimental	
				experimental	
				in silico	
	146.0691			*in silico*	

https://metlin.scripps.edu/showChart.php?molid=18&h=240&collE=20&lmode=p&etype=experimental

L-Glutamine

ID: 18 ESI Q-TOF

Mode: (+) Collision Energy: 40 V

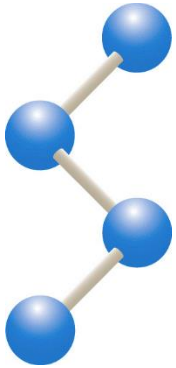
m/z: 56.0502 Intensity: 100 %

🟦	(+)	0 V	[M+H]+
🟥	(+)	10 V	[M+H]+
🟩	(+)	20 V	[M+H]+
🟪	(+)	40 V	[M+H]+
🟦	(-)	0 V	[M-H]-
🟧	(-)	10 V	[M-H]-
🟦	(-)	20 V	[M-H]-
🟫	(-)	40 V	[M-H]-

Predicted Fragment Structure [M]+, Mass: 56.0500

⚠ Please mouse over the spectrum to view the detail information of each peak
Use left mouse button to zoom in (click and drag) and zoom out (double-click)

May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Statistical Analysis***

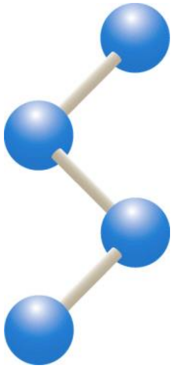
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

May 16th



Fundamental Metabolomics

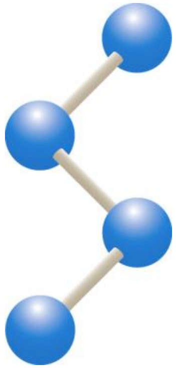
- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Targeted Metabolomics***
- ***Untargeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----



Fundamental Metabolomics

May 16th

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Targeted Metabolomics***
- ***Untargeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

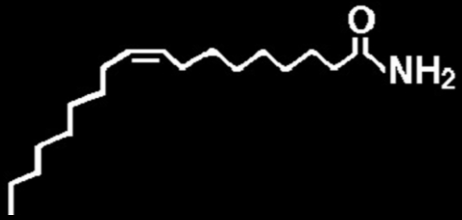
May 16th

---- 12:30 pm Begin ----

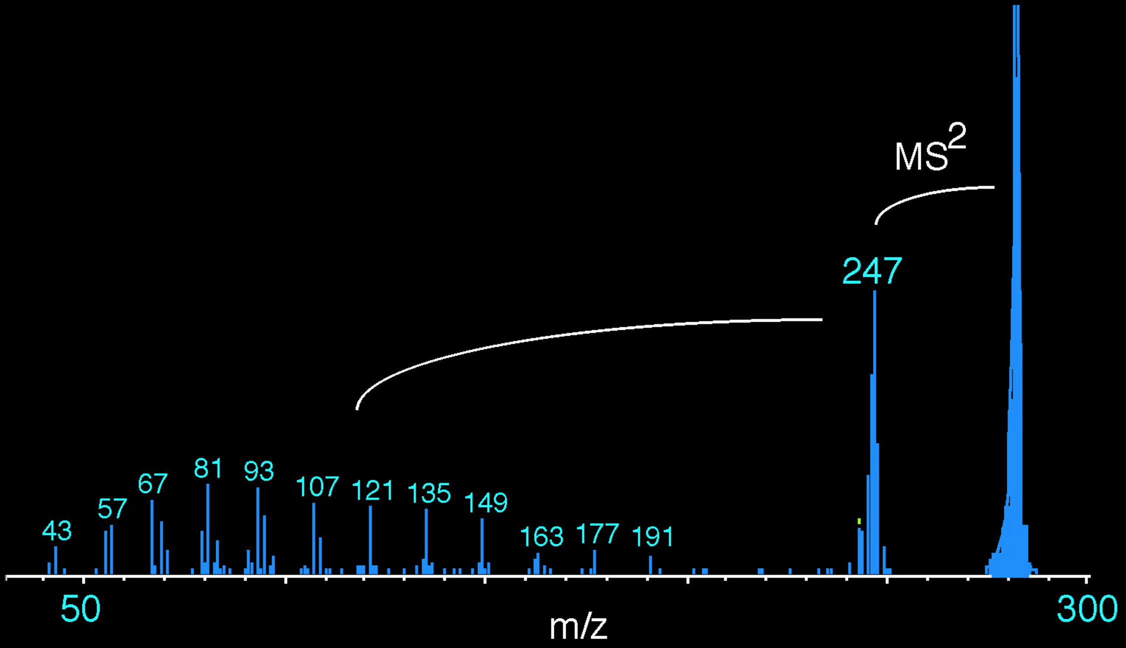
---- 3:00 pm Break ----

---- 04:30 pm Finish ----

Sleep



- Tandem MS
- Accurate m/z
- H/D Exchange
- Isotope Pattern



- Preparative LC for collection of metabolite for NMR
- Elemental composition of MS^2 and MS^3 fragment ions
- Synthesize authentic standard

Science 1995

Technology

Science 1995

Technology

It took about 8 months to identify oleamide

Science 1995

Technology

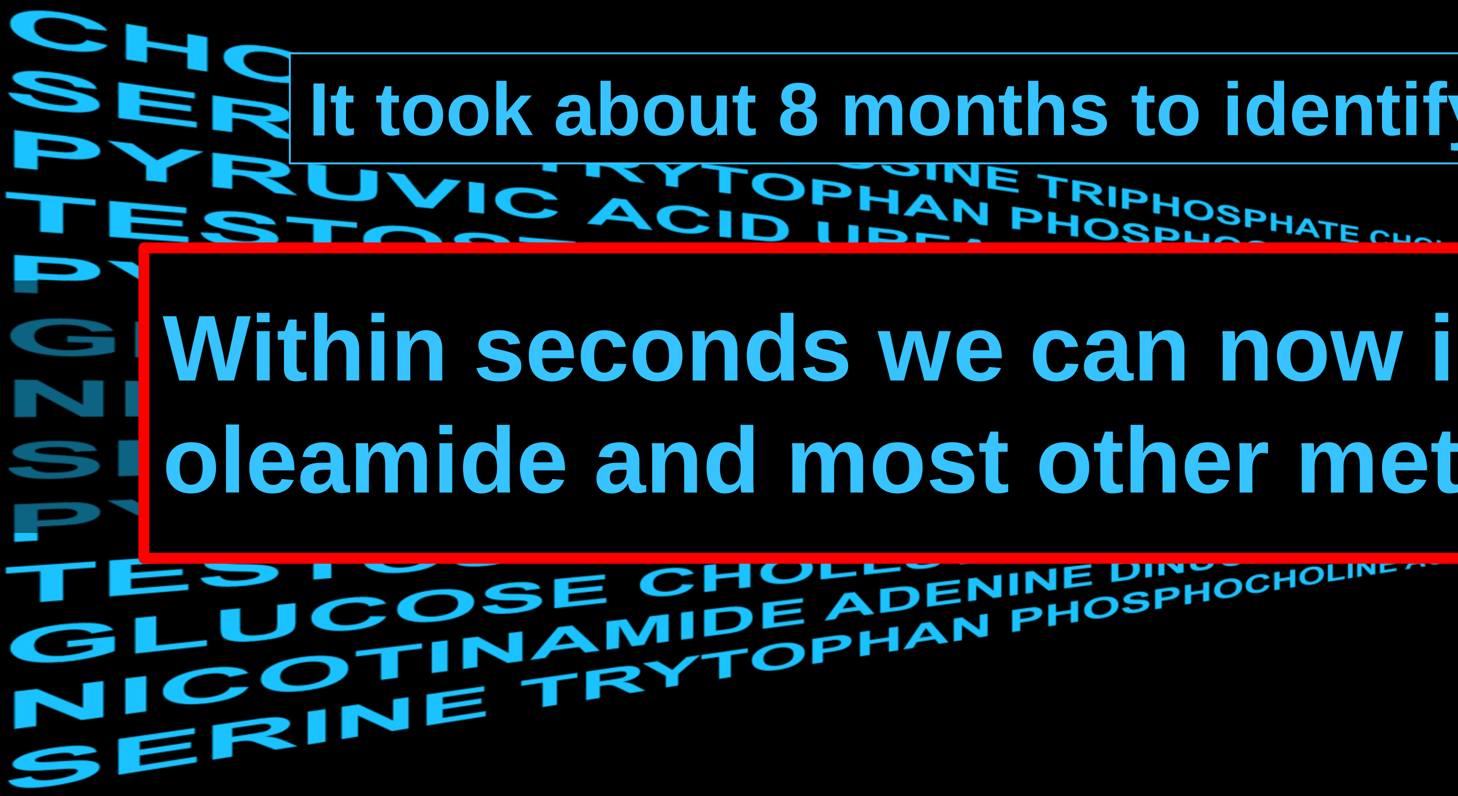
It took about 8 months to identify oleamide

METLIN

1 Million Molecules in METLIN

Science 1995

Technology

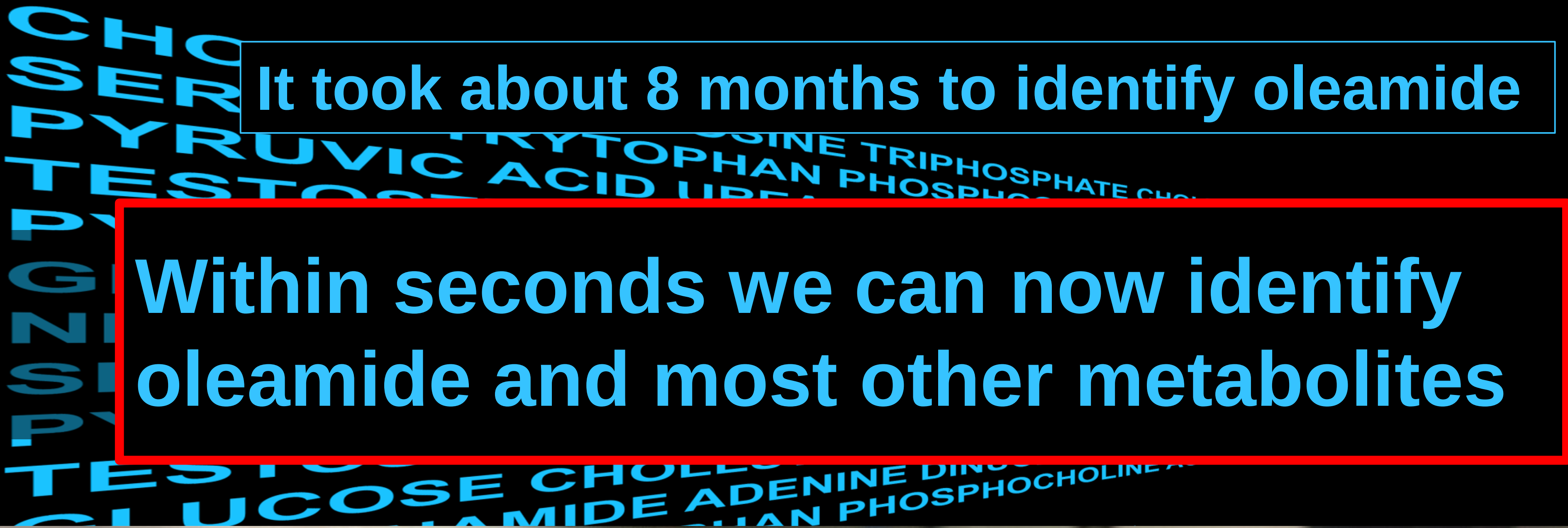


It took about 8 months to identify oleamide

Within seconds we can now identify oleamide and most other metabolites

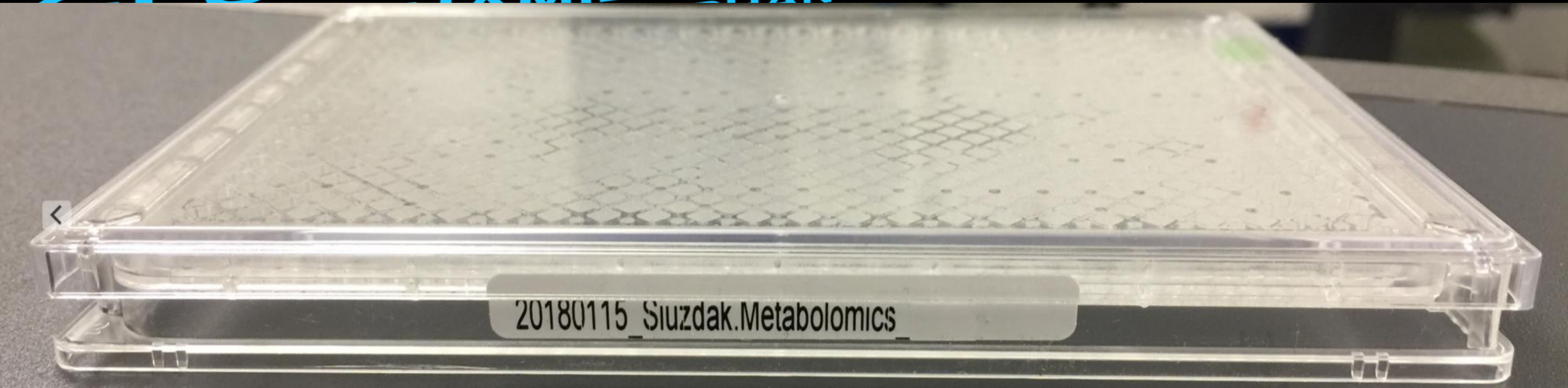
1 Million Molecules in METLIN

Technology



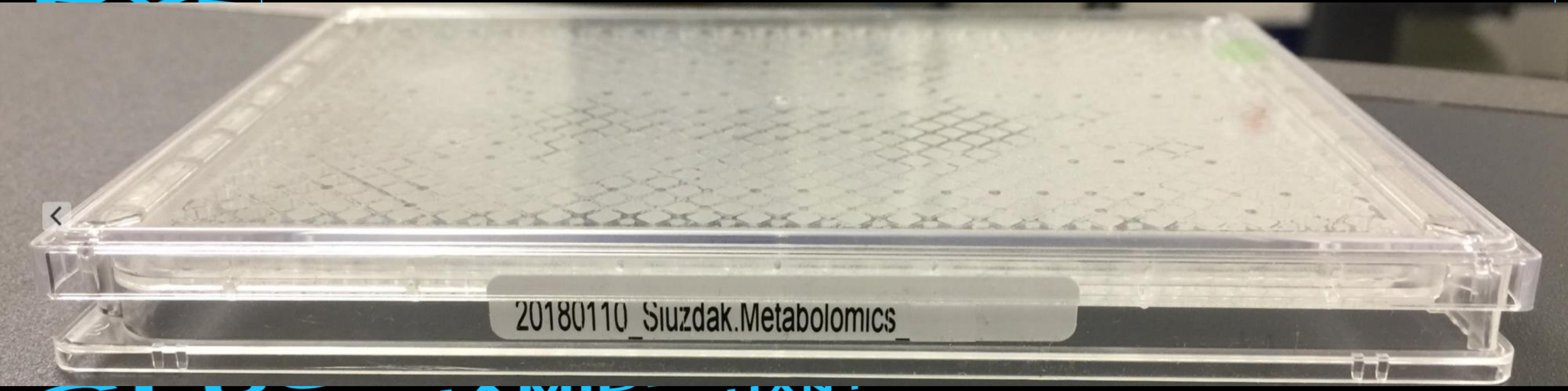
It took about 8 months to identify oleamide

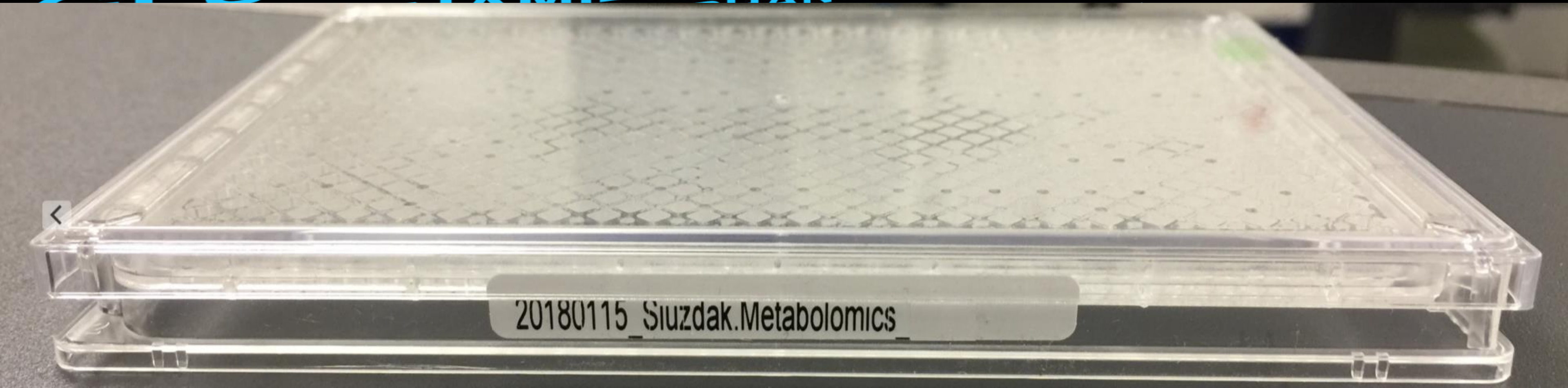
Within seconds we can now identify oleamide and most other metabolites

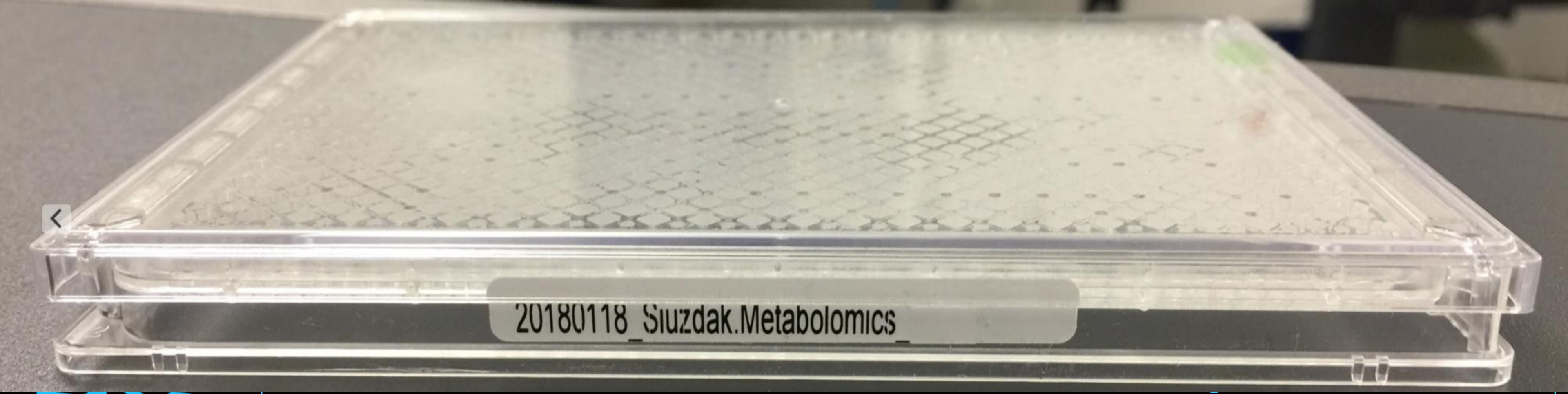


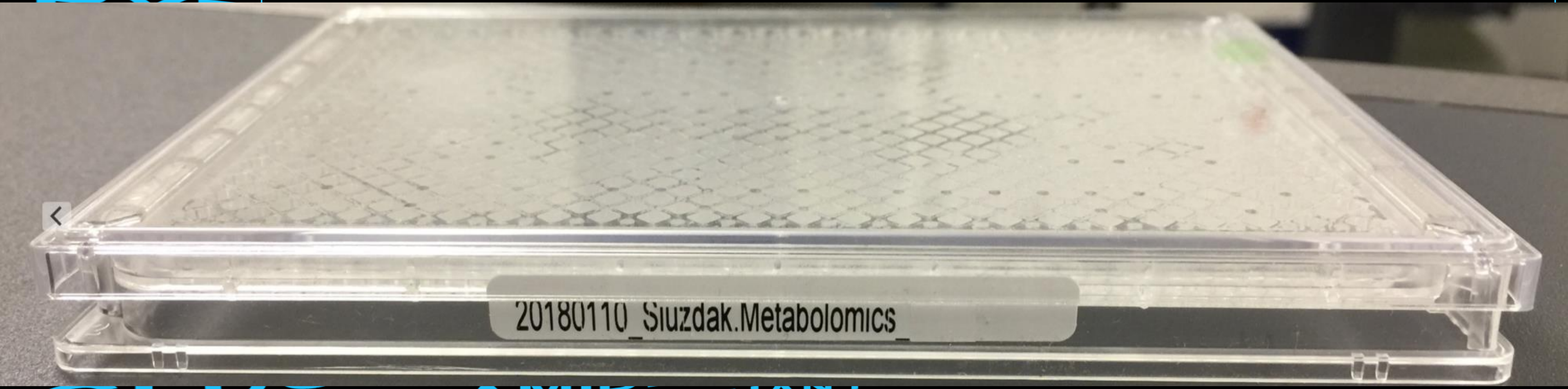
Technology

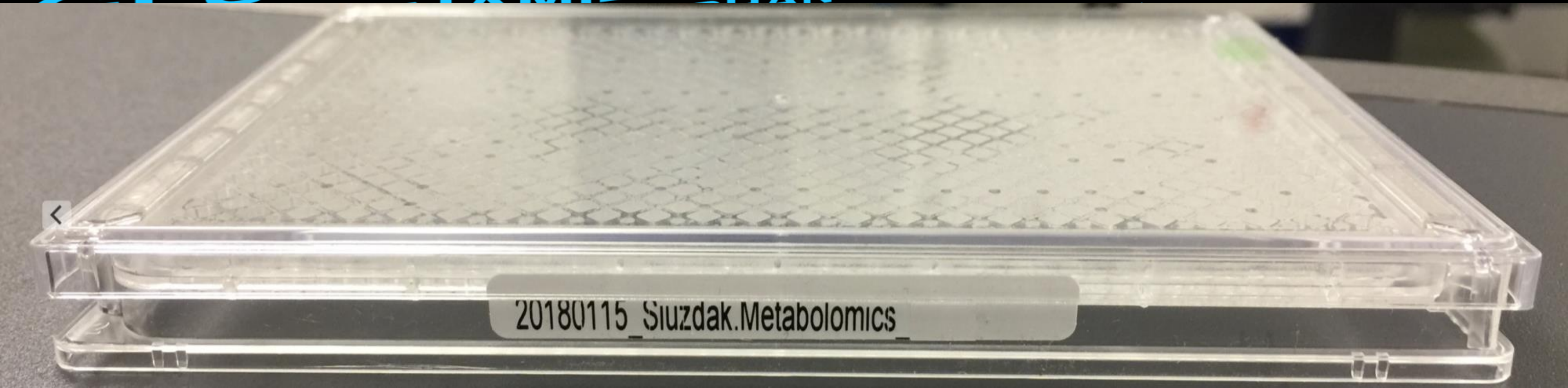
It took about 8 months to identify oleamide






20180118_Siuzdak.Metabolomics_


20180110_Siuzdak.Metabolomics_


20180115_Siuzdak.Metabolomics_

METLIN

1 Million Molecules in METLIN

15,000 Standards with MS/MS data

1 Million Molecules in METLIN

15,000 Standards with MS/MS data

METLIN

1 Million Molecules in METLIN

60,000 Standards with MS/MS data

The Standards

1 Million Molecules in METLIN

60,000 Standards with MS/MS data

The Standards

Sample Preparation

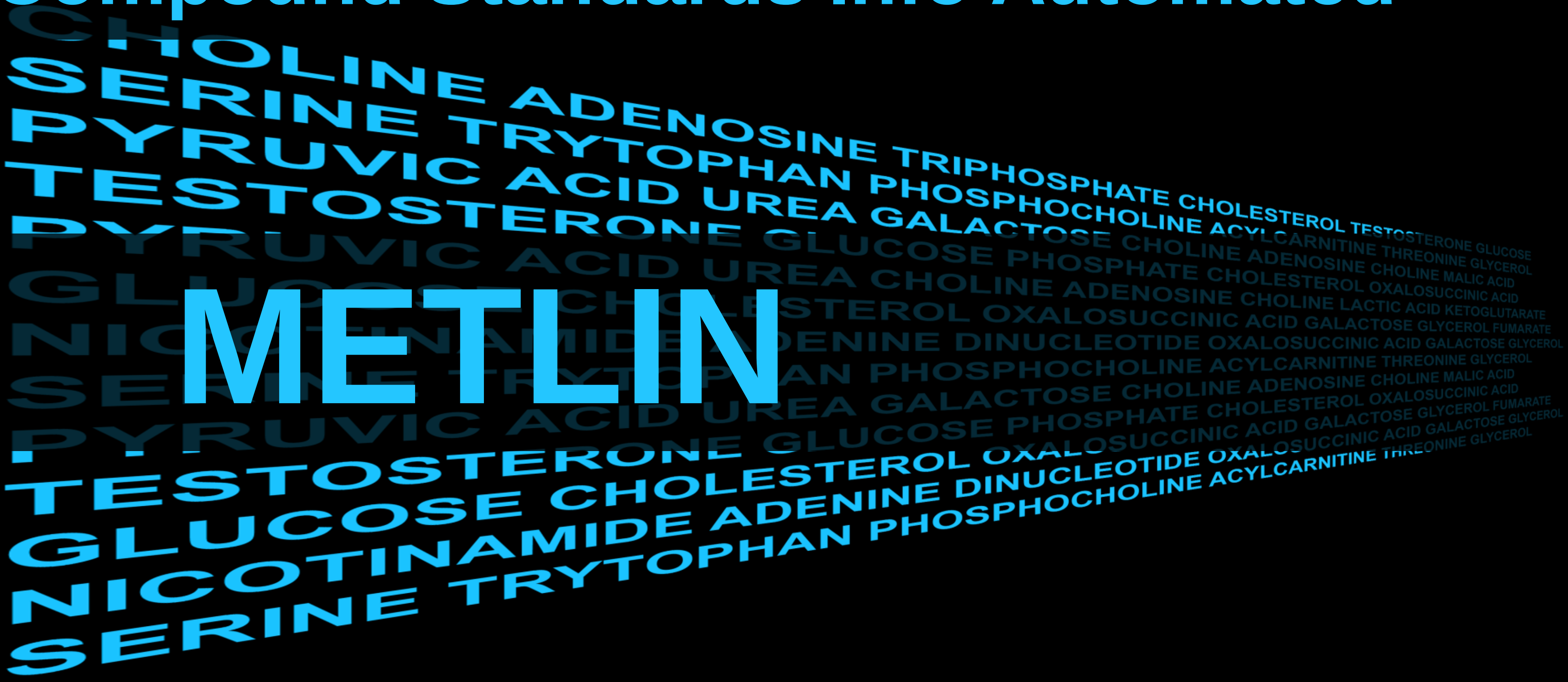
1 Million Molecules in METLIN

60,000 Standards with MS/MS data

The Standards

Sample Preparation

Compound Standards Info Automated



1 Million Molecules in METLIN

60,000 Standards with MS/MS data

The Standards

Sample Preparation

Compound Standards Info Automated

Reliable & Accurate MS Data Acquisition

1 Million Molecules in METLIN

60,000 Standards with MS/MS data

The Standards

Sample Preparation

Compound Standards Info Automated

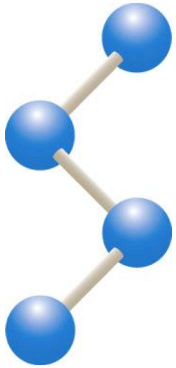
Reliable & Accurate MS Data Acquisition

Data Streamed into METLIN Autonomously

METLIN

1 Million Molecules in METLIN

60,000 Standards with MS/MS data

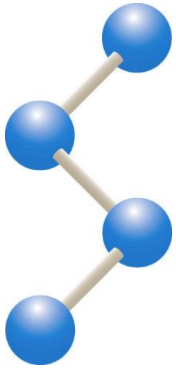


May 16th

Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Targeted Metabolomics***
- ***Untargeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

May 16th
---- 12:30 pm Begin ----
---- 3:00 pm Break ----
---- 04:30 pm Finish ----

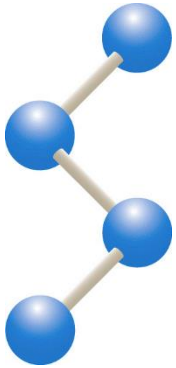


Fundamental Metabolomics

May 16th

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

May 16th
---- 12:30 pm Begin ----
---- 3:00 pm Break ----
---- 04:30 pm Finish ----



Fundamental Metabolomics

May 16th

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

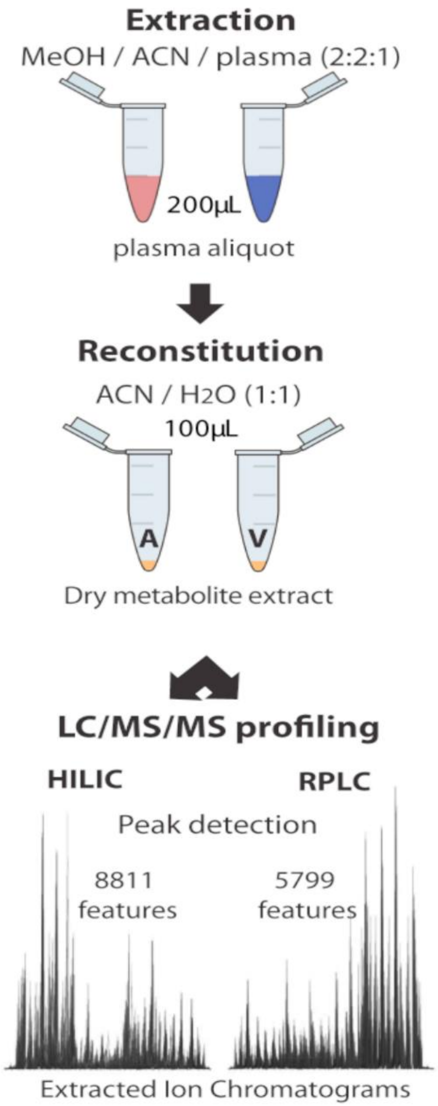
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

Untargeted Metabolomics



Extraction

MeOH / ACN / plasma (2:2:1)

200μL plasma aliquot

Reconstitution

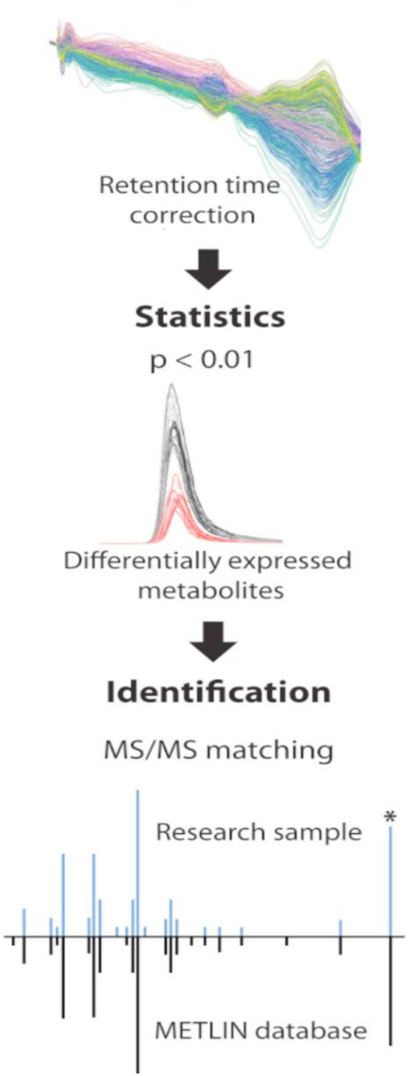
ACN / H_2O (1:1)

100μL

Dry metabolite extract

LC/MS/MS profiling

Data Processing



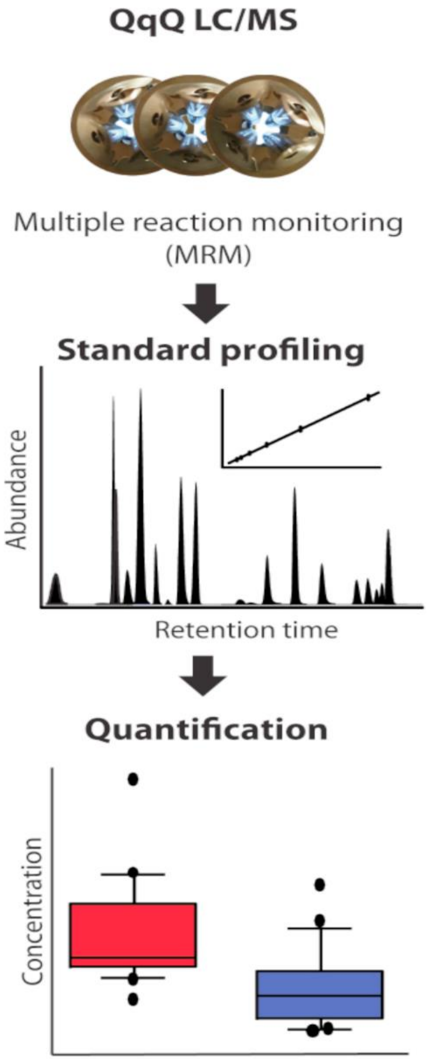
Statistics

$p < 0.01$

Identification

MS/MS matching

Targeted Metabolomics



QqQ LC/MS

Multiple reaction monitoring (MRM)

Standard profiling

Quantification

May 16th

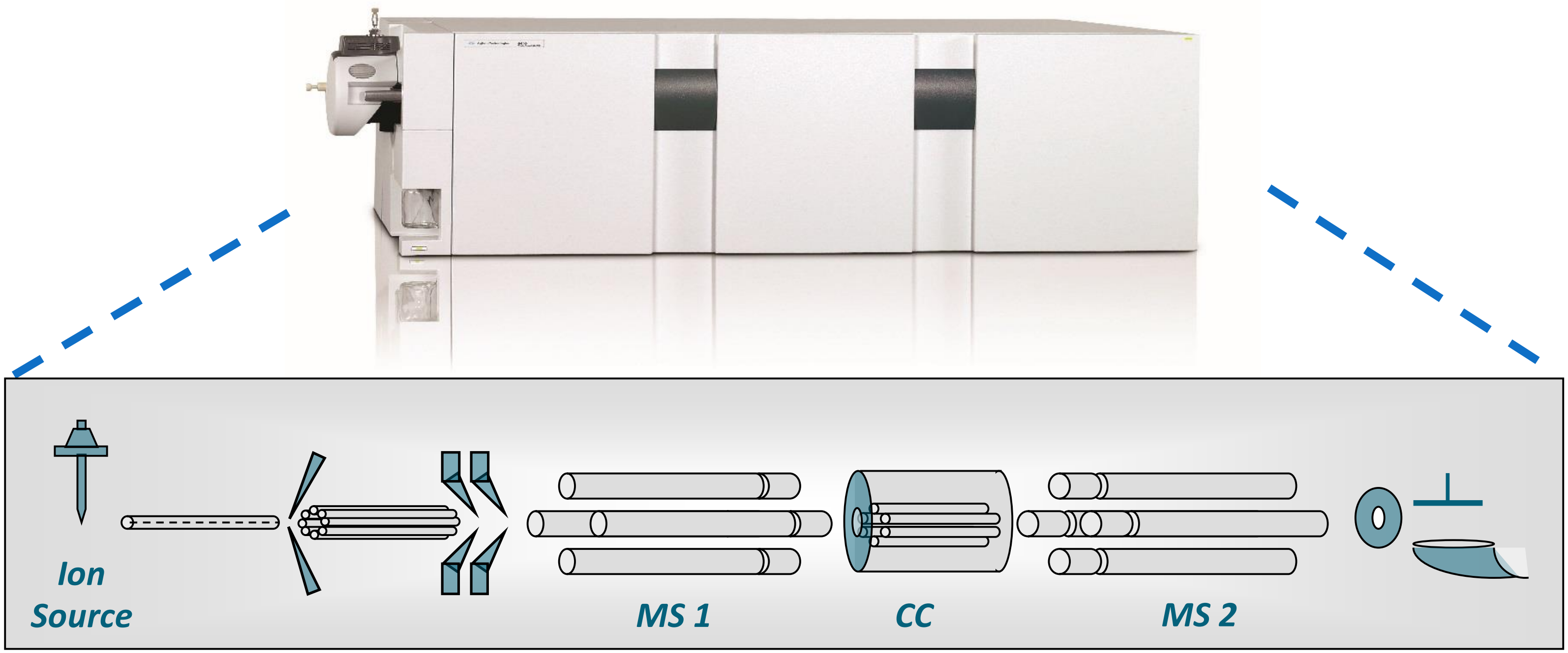
Fundamental Metabolomics

- Targeted metabolomics
- Know which compounds to measure
- Concentration

May 16th

Fundamental Metabolomics

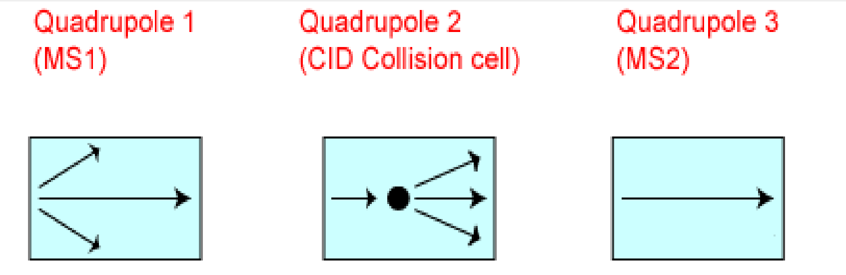
Triple Quadrupole Mass Spectrometers



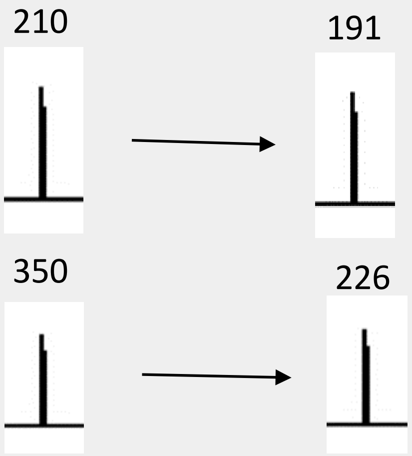
May 16th

Fundamental Metabolomics

A mass selected precursor ion is chosen with Q1 and the collision cell generates fragment ions. Q3 is set to look for specific fragment ions. This is known as multiple reaction monitoring, or MRM.



Q1 lets **only** target ion 210 and 350 pass through

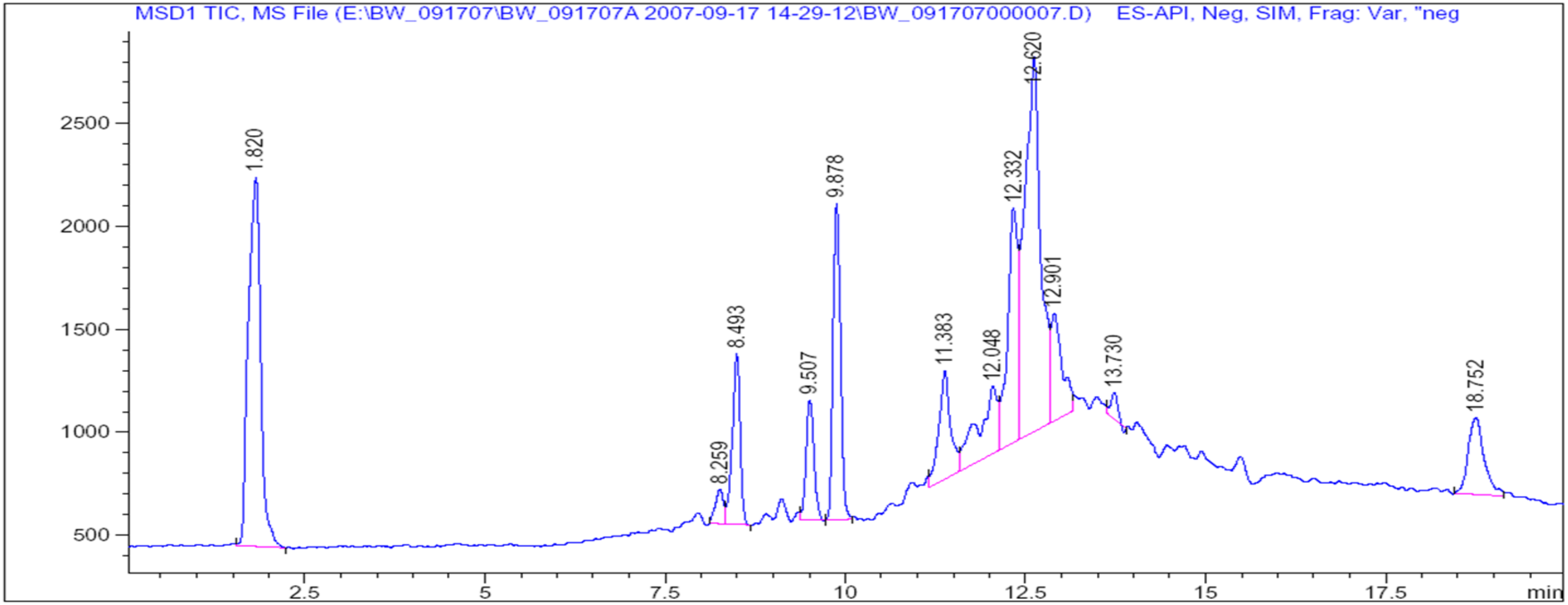


Q3 monitors **only** characteristic fragments 191 from ion 210 and fragment 226 from target ion 350 for quantitative analysis

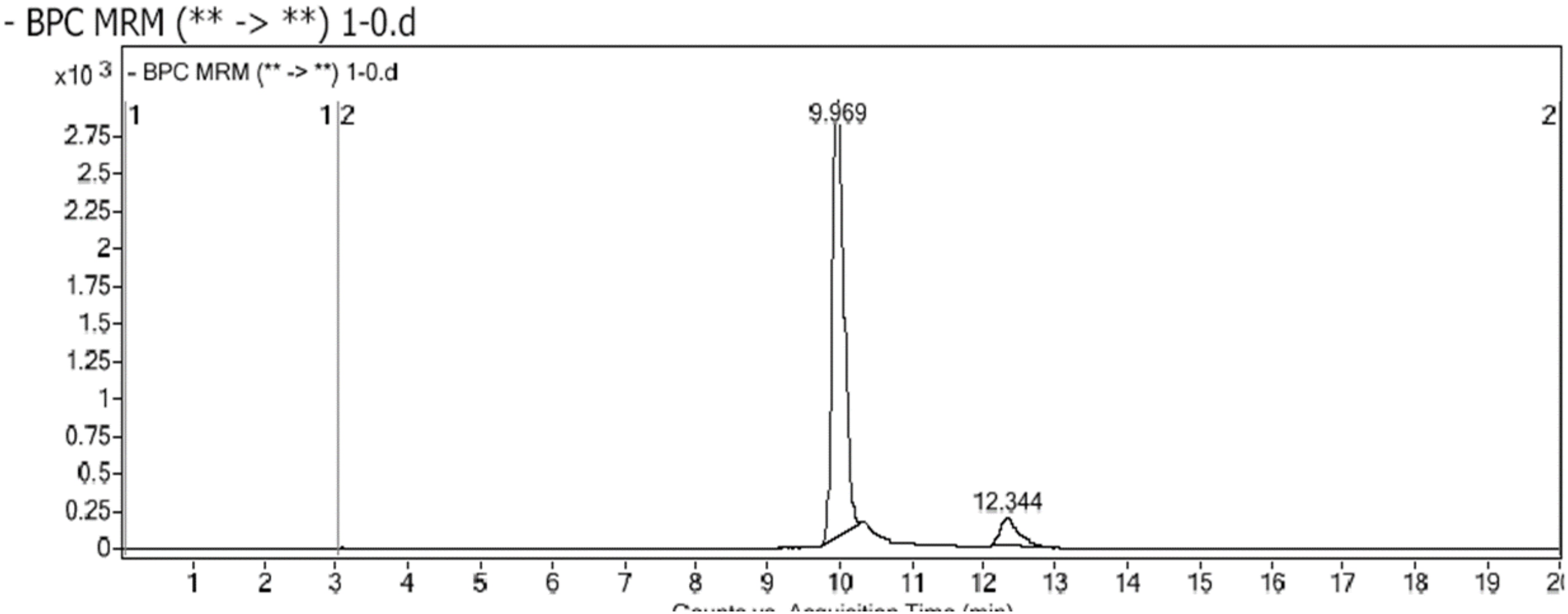
May 16th

Fundamental Metabolomics

Single quad in Selected Ion Monitoring mode



Triple Quad in MRM mode
Specificity and sensitivity



Fundamental Metabolomics

Internal standards

- Internal standards take into account losses during prep

Fundamental Metabolomics

Internal standards

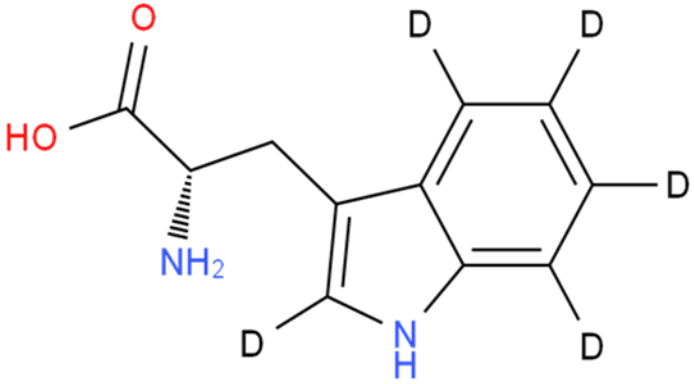
- Internal standards take into account losses during prep
- Will adjust for matrix effects and suppression

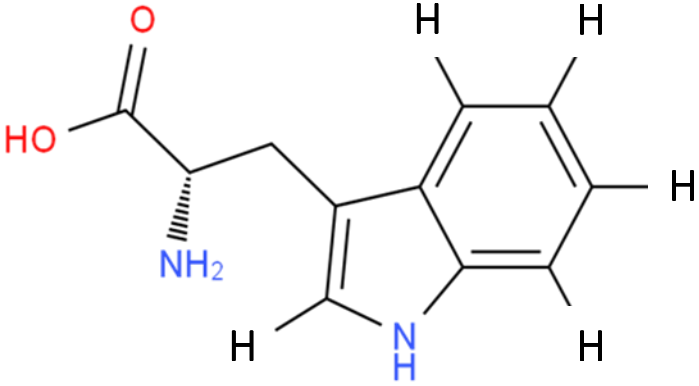
Internal standards

- Internal standards take into account losses during prep
- Will adjust for matrix effects and suppression
- Chemically similar, but differ in mass

Internal standards

- Internal standards take into account losses during prep
- Will adjust for matrix effects and suppression
- Chemically similar, but differ in mass



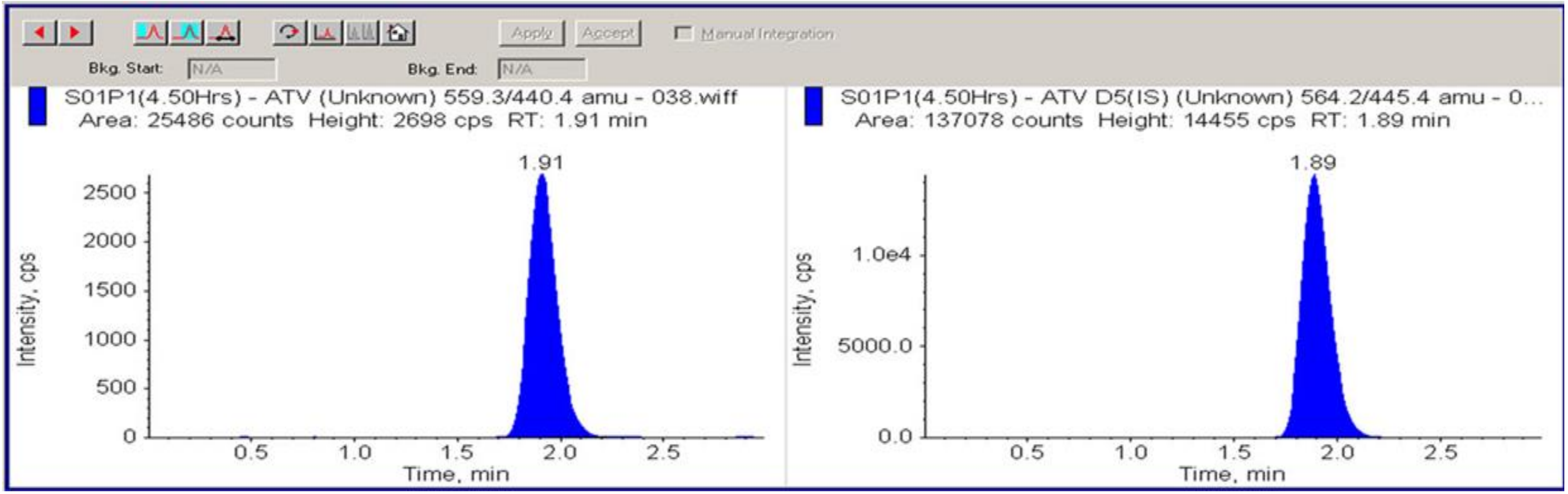


May 16th

Fundamental Metabolomics

Internal standards

- Internal standards are added to samples and standards at the same concentration
- Absolute response will differ due to losses, but ratio of response will remain the same



Bkg. Start: N/A

Bkg. End: N/A

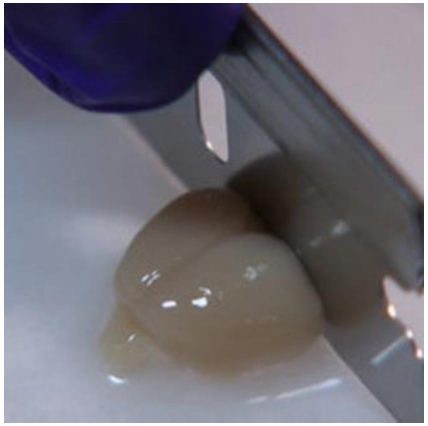
S01P1(4.50Hrs) - ATV (Unknown) 559.3/440.4 amu - 038.wiff
Area: 25486 counts Height: 2698 cps RT: 1.91 min

S01P1(4.50Hrs) - ATV D5(IS) (Unknown) 564.2/445.4 amu - 0...
Area: 137078 counts Height: 14455 cps RT: 1.89 min

Fundamental Metabolomics

Internal standards
Addition to samples

Internal standards are
At the same concentration
In samples and standards



May 16th

Fundamental Metabolomics

Internal standards
Addition to samples

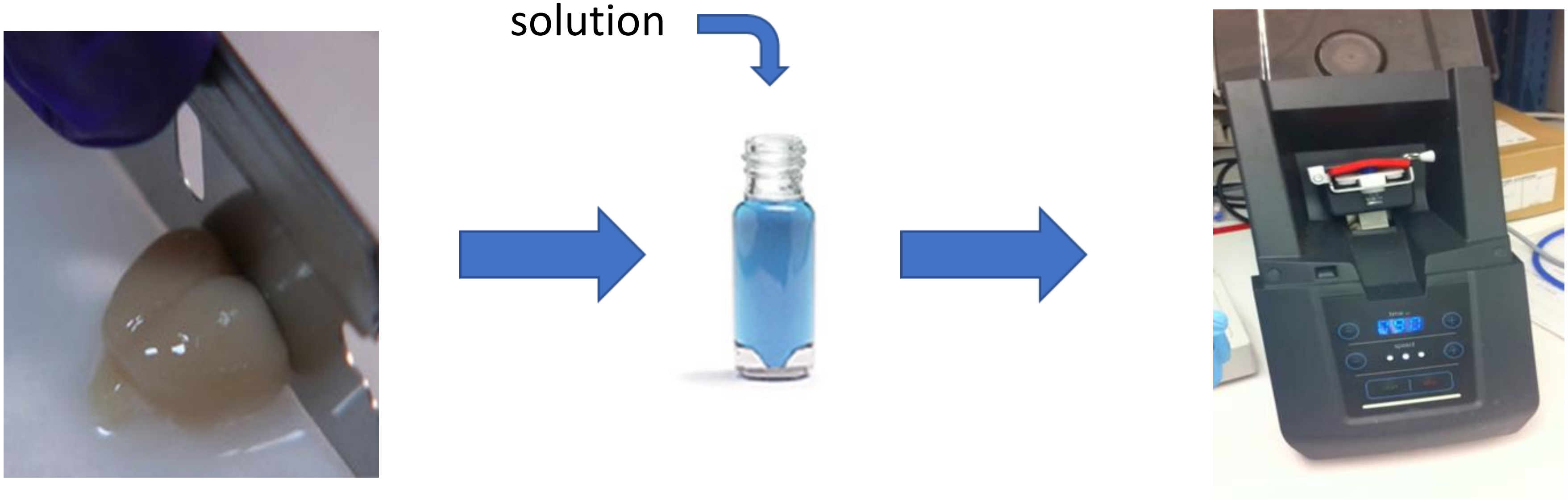
For liquid samples – plasma, urine, csf spike directly into vial prior to adding extraction solvent

Fundamental Metabolomics

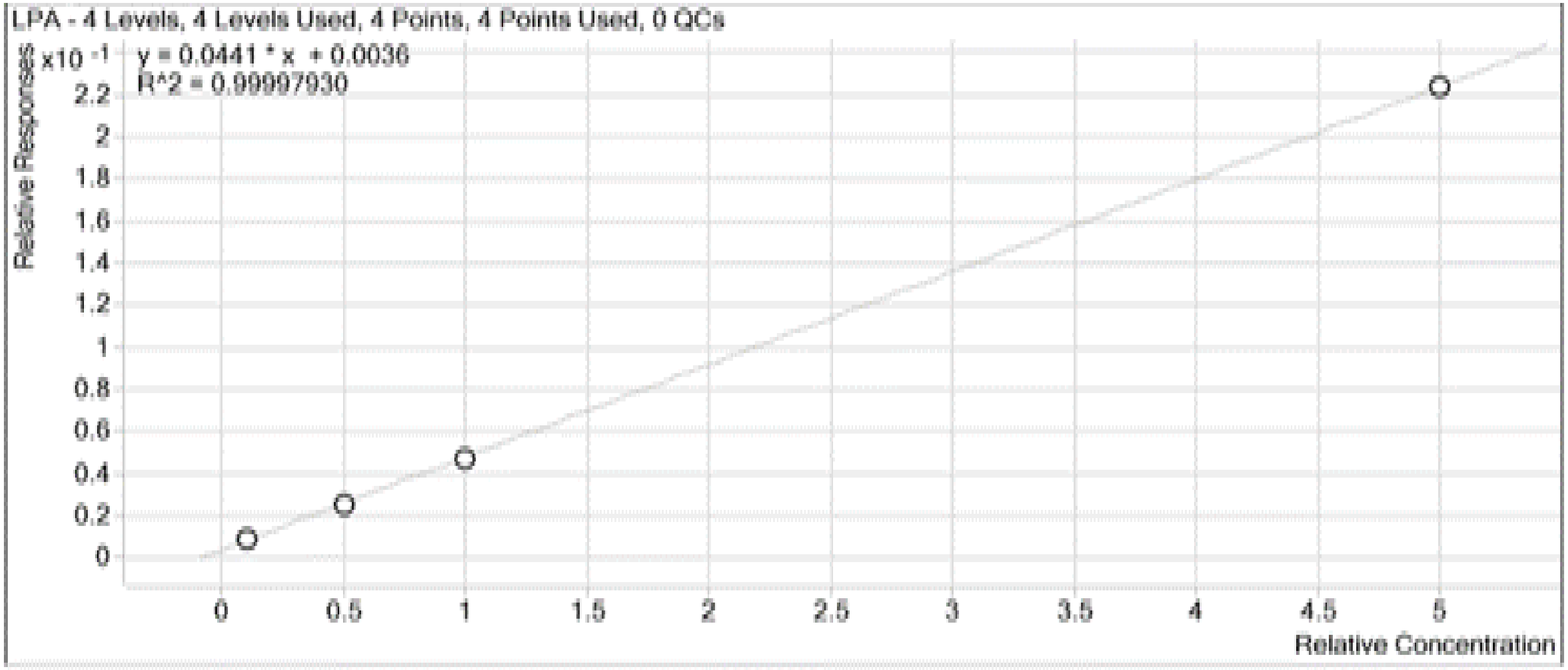
Internal standards

Addition to samples

For tissue samples
add I.S. to extraction
solution

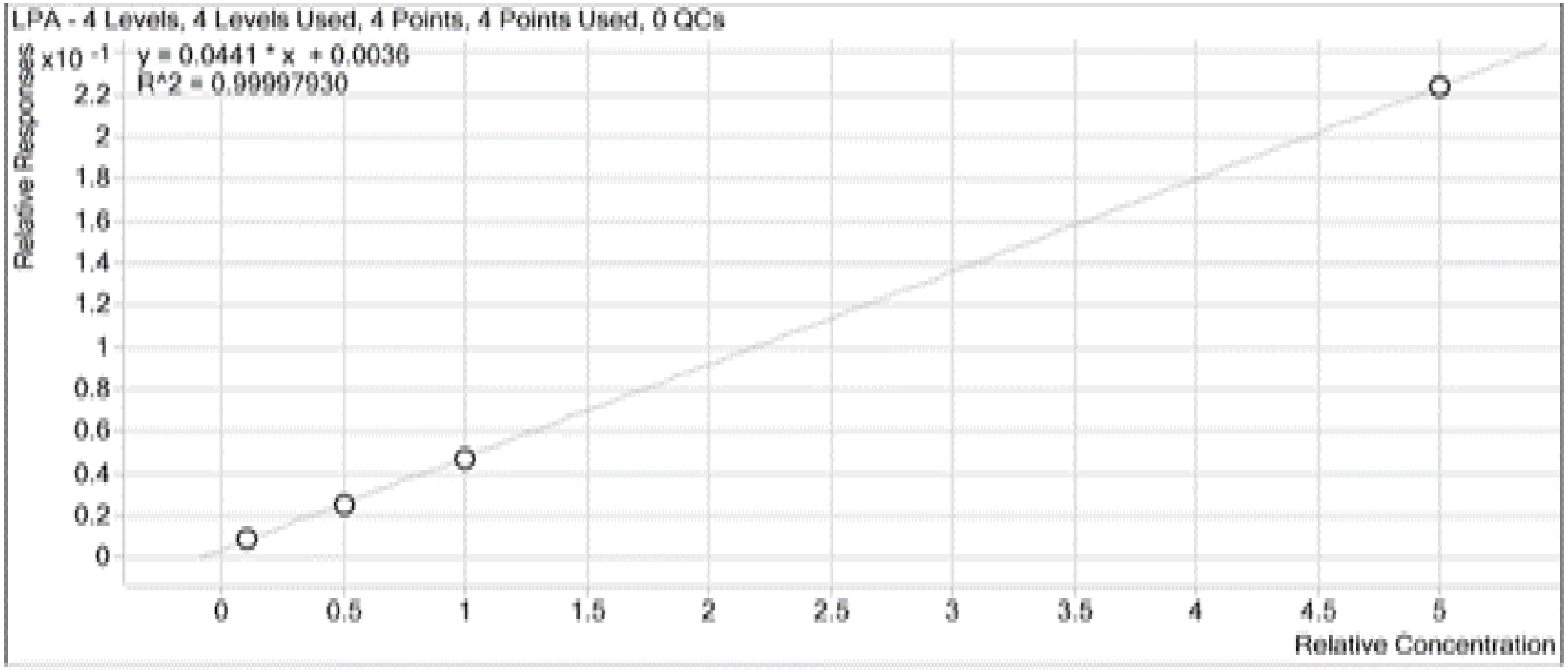


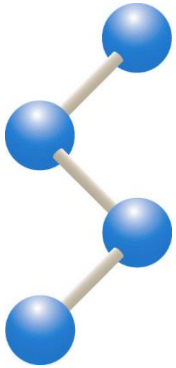
Results in a linear standard curve



Fundamental Metabolomics

Results in a linear standard curve
Concentrations should bracket what is expected
In the samples





May 16th

Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

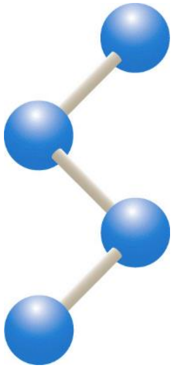
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

May 16th



Fundamental Metabolomics

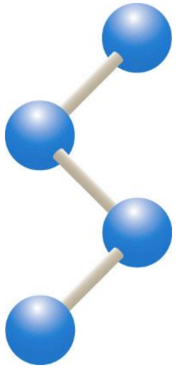
- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Statistical Analysis***

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----



Fundamental Metabolomics

May 16th

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Statistical Analysis***

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

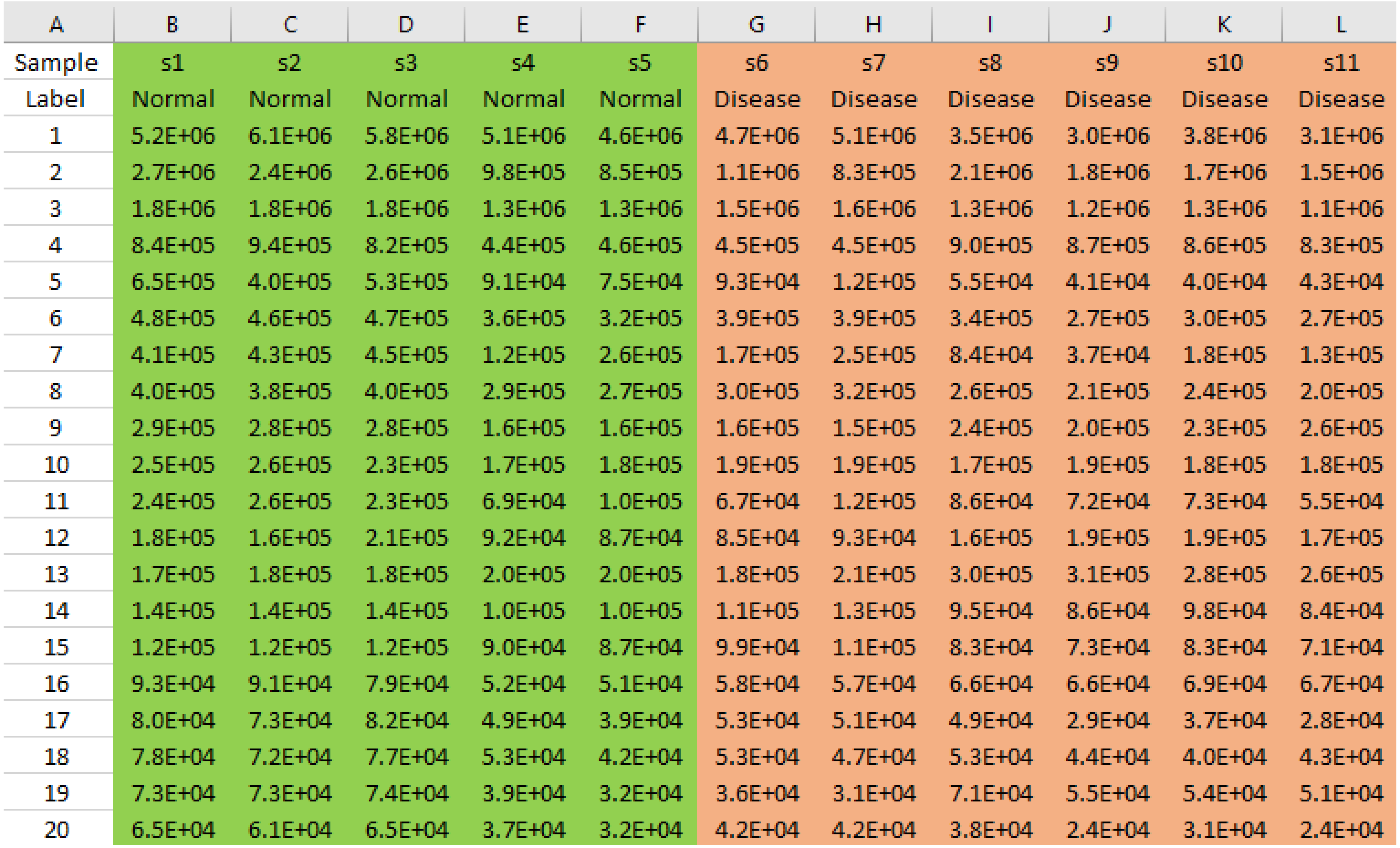
---- 04:30 pm Finish ----

Fundamental Metabolomics

Statistics is the mathematics of impression

Fundamental Metabolomics

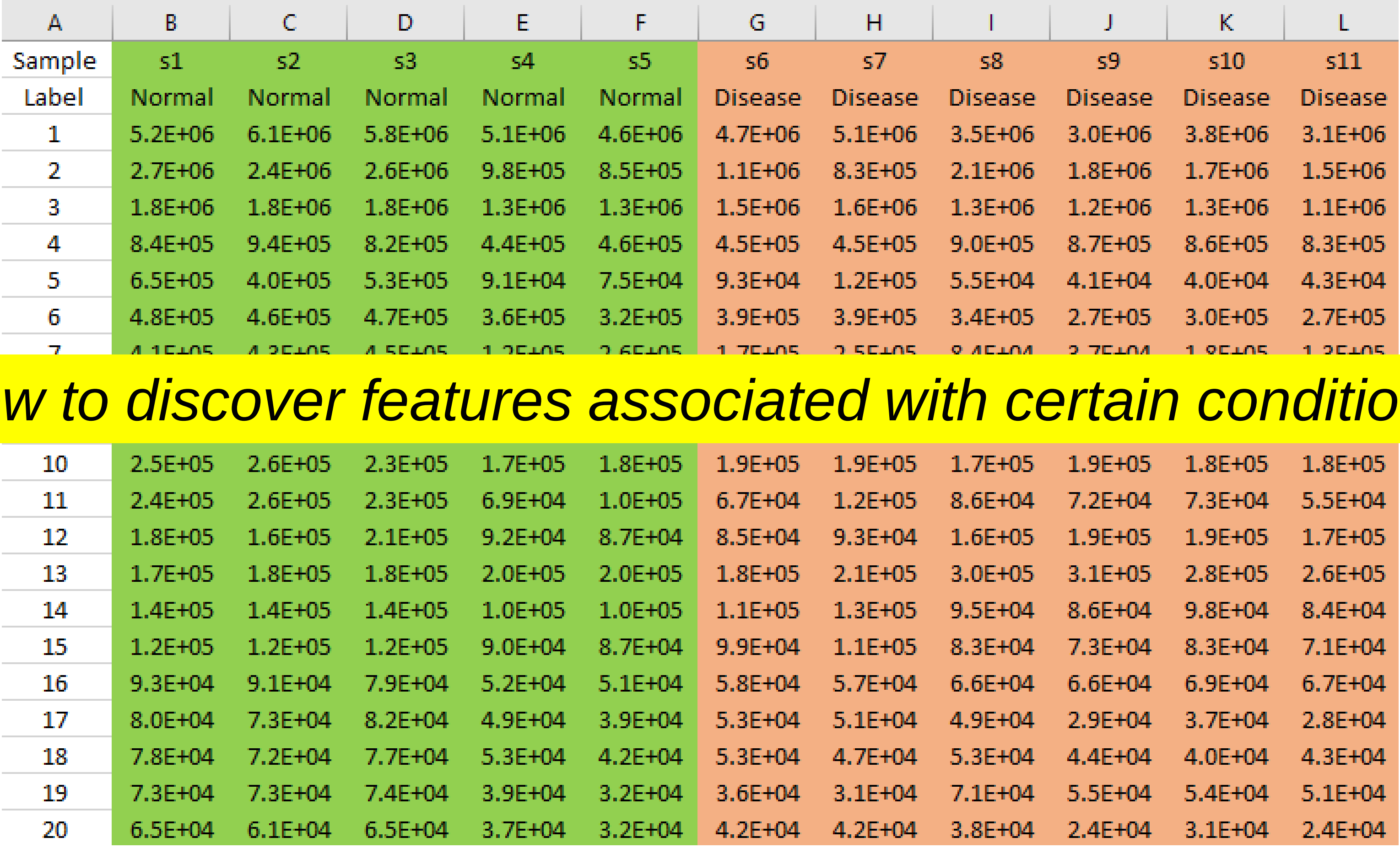
Statistics is the mathematics of impression



A	B	C	D	E	F	G	H	I	J	K	L
Sample	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11
Label	Normal	Normal	Normal	Normal	Normal	Disease	Disease	Disease	Disease	Disease	Disease
1	5.2E+06	6.1E+06	5.8E+06	5.1E+06	4.6E+06	4.7E+06	5.1E+06	3.5E+06	3.0E+06	3.8E+06	3.1E+06
2	2.7E+06	2.4E+06	2.6E+06	9.8E+05	8.5E+05	1.1E+06	8.3E+05	2.1E+06	1.8E+06	1.7E+06	1.5E+06
3	1.8E+06	1.8E+06	1.8E+06	1.3E+06	1.3E+06	1.5E+06	1.6E+06	1.3E+06	1.2E+06	1.3E+06	1.1E+06
4	8.4E+05	9.4E+05	8.2E+05	4.4E+05	4.6E+05	4.5E+05	4.5E+05	9.0E+05	8.7E+05	8.6E+05	8.3E+05
5	6.5E+05	4.0E+05	5.3E+05	9.1E+04	7.5E+04	9.3E+04	1.2E+05	5.5E+04	4.1E+04	4.0E+04	4.3E+04
6	4.8E+05	4.6E+05	4.7E+05	3.6E+05	3.2E+05	3.9E+05	3.9E+05	3.4E+05	2.7E+05	3.0E+05	2.7E+05
7	4.1E+05	4.3E+05	4.5E+05	1.2E+05	2.6E+05	1.7E+05	2.5E+05	8.4E+04	3.7E+04	1.8E+05	1.3E+05
8	4.0E+05	3.8E+05	4.0E+05	2.9E+05	2.7E+05	3.0E+05	3.2E+05	2.6E+05	2.1E+05	2.4E+05	2.0E+05
9	2.9E+05	2.8E+05	2.8E+05	1.6E+05	1.6E+05	1.6E+05	1.5E+05	2.4E+05	2.0E+05	2.3E+05	2.6E+05
10	2.5E+05	2.6E+05	2.3E+05	1.7E+05	1.8E+05	1.9E+05	1.9E+05	1.7E+05	1.9E+05	1.8E+05	1.8E+05
11	2.4E+05	2.6E+05	2.3E+05	6.9E+04	1.0E+05	6.7E+04	1.2E+05	8.6E+04	7.2E+04	7.3E+04	5.5E+04
12	1.8E+05	1.6E+05	2.1E+05	9.2E+04	8.7E+04	8.5E+04	9.3E+04	1.6E+05	1.9E+05	1.9E+05	1.7E+05
13	1.7E+05	1.8E+05	1.8E+05	2.0E+05	2.0E+05	1.8E+05	2.1E+05	3.0E+05	3.1E+05	2.8E+05	2.6E+05
14	1.4E+05	1.4E+05	1.4E+05	1.0E+05	1.0E+05	1.1E+05	1.3E+05	9.5E+04	8.6E+04	9.8E+04	8.4E+04
15	1.2E+05	1.2E+05	1.2E+05	9.0E+04	8.7E+04	9.9E+04	1.1E+05	8.3E+04	7.3E+04	8.3E+04	7.1E+04
16	9.3E+04	9.1E+04	7.9E+04	5.2E+04	5.1E+04	5.8E+04	5.7E+04	6.6E+04	6.6E+04	6.9E+04	6.7E+04
17	8.0E+04	7.3E+04	8.2E+04	4.9E+04	3.9E+04	5.3E+04	5.1E+04	4.9E+04	2.9E+04	3.7E+04	2.8E+04
18	7.8E+04	7.2E+04	7.7E+04	5.3E+04	4.2E+04	5.3E+04	4.7E+04	5.3E+04	4.4E+04	4.0E+04	4.3E+04
19	7.3E+04	7.3E+04	7.4E+04	3.9E+04	3.2E+04	3.6E+04	3.1E+04	7.1E+04	5.5E+04	5.4E+04	5.1E+04
20	6.5E+04	6.1E+04	6.5E+04	3.7E+04	3.2E+04	4.2E+04	4.2E+04	3.8E+04	2.4E+04	3.1E+04	2.4E+04

Fundamental Metabolomics

Statistics is the mathematics of impression



A	B	C	D	E	F	G	H	I	J	K	L
Sample	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11
Label	Normal	Normal	Normal	Normal	Normal	Disease	Disease	Disease	Disease	Disease	Disease
1	5.2E+06	6.1E+06	5.8E+06	5.1E+06	4.6E+06	4.7E+06	5.1E+06	3.5E+06	3.0E+06	3.8E+06	3.1E+06
2	2.7E+06	2.4E+06	2.6E+06	9.8E+05	8.5E+05	1.1E+06	8.3E+05	2.1E+06	1.8E+06	1.7E+06	1.5E+06
3	1.8E+06	1.8E+06	1.8E+06	1.3E+06	1.3E+06	1.5E+06	1.6E+06	1.3E+06	1.2E+06	1.3E+06	1.1E+06
4	8.4E+05	9.4E+05	8.2E+05	4.4E+05	4.6E+05	4.5E+05	4.5E+05	9.0E+05	8.7E+05	8.6E+05	8.3E+05
5	6.5E+05	4.0E+05	5.3E+05	9.1E+04	7.5E+04	9.3E+04	1.2E+05	5.5E+04	4.1E+04	4.0E+04	4.3E+04
6	4.8E+05	4.6E+05	4.7E+05	3.6E+05	3.2E+05	3.9E+05	3.9E+05	3.4E+05	2.7E+05	3.0E+05	2.7E+05
7	4.1E+05	4.3E+05	4.5E+05	1.2E+05	2.6E+05	1.7E+05	2.5E+05	8.4E+04	3.7E+04	1.8E+05	1.3E+05
10	2.5E+05	2.6E+05	2.3E+05	1.7E+05	1.8E+05	1.9E+05	1.9E+05	1.7E+05	1.9E+05	1.8E+05	1.8E+05
11	2.4E+05	2.6E+05	2.3E+05	6.9E+04	1.0E+05	6.7E+04	1.2E+05	8.6E+04	7.2E+04	7.3E+04	5.5E+04
12	1.8E+05	1.6E+05	2.1E+05	9.2E+04	8.7E+04	8.5E+04	9.3E+04	1.6E+05	1.9E+05	1.9E+05	1.7E+05
13	1.7E+05	1.8E+05	1.8E+05	2.0E+05	2.0E+05	1.8E+05	2.1E+05	3.0E+05	3.1E+05	2.8E+05	2.6E+05
14	1.4E+05	1.4E+05	1.4E+05	1.0E+05	1.0E+05	1.1E+05	1.3E+05	9.5E+04	8.6E+04	9.8E+04	8.4E+04
15	1.2E+05	1.2E+05	1.2E+05	9.0E+04	8.7E+04	9.9E+04	1.1E+05	8.3E+04	7.3E+04	8.3E+04	7.1E+04
16	9.3E+04	9.1E+04	7.9E+04	5.2E+04	5.1E+04	5.8E+04	5.7E+04	6.6E+04	6.6E+04	6.9E+04	6.7E+04
17	8.0E+04	7.3E+04	8.2E+04	4.9E+04	3.9E+04	5.3E+04	5.1E+04	4.9E+04	2.9E+04	3.7E+04	2.8E+04
18	7.8E+04	7.2E+04	7.7E+04	5.3E+04	4.2E+04	5.3E+04	4.7E+04	5.3E+04	4.4E+04	4.0E+04	4.3E+04
19	7.3E+04	7.3E+04	7.4E+04	3.9E+04	3.2E+04	3.6E+04	3.1E+04	7.1E+04	5.5E+04	5.4E+04	5.1E+04
20	6.5E+04	6.1E+04	6.5E+04	3.7E+04	3.2E+04	4.2E+04	4.2E+04	3.8E+04	2.4E+04	3.1E+04	2.4E+04

How to discover features associated with certain condition?

Fundamental Metabolomics

- Univariate analysis examines each variable separately, providing a ranking of potentially important features
  - ✓ T-test
  - ✓ Volcano plot

Fundamental Metabolomics

- Univariate analysis examines each variable separately, providing a ranking of potentially important features
  - ✓ T-test
  - ✓ Volcano plot

- Multivariate analysis considers two or more variables simultaneously and takes into count relationships between variables
  - ✓ PCA: Principle Component Analysis
  - ✓ PLC-DA: Partial Least Squares-Discriminant Analysis

Fundamental Metabolomics

Univariate analysis: Student’s t-Test

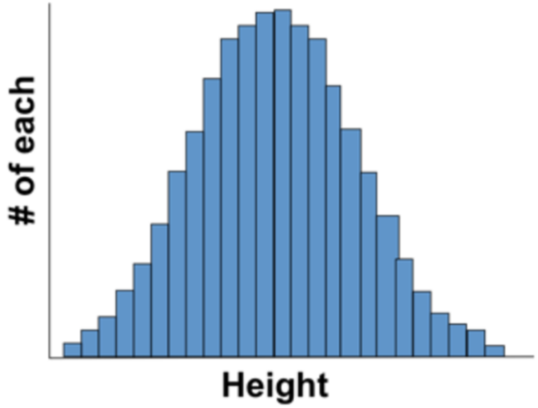
- Also called the t-test
- Used to determine if 2 populations are different

Data distribution

A basic way of presenting univariate data is to create a frequency distribution of the individual cases

Data distribution

A basic way of presenting univariate data is to create a frequency distribution of the individual cases

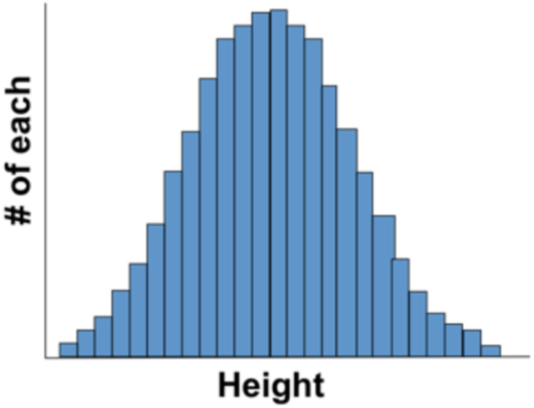


Fundamental Metabolomics

Data distribution

A basic way of presenting univariate data is to create a frequency distribution of the individual cases

Frequency distribution can be modeled as a normal/Gaussian distribution



Fundamental Metabolomics

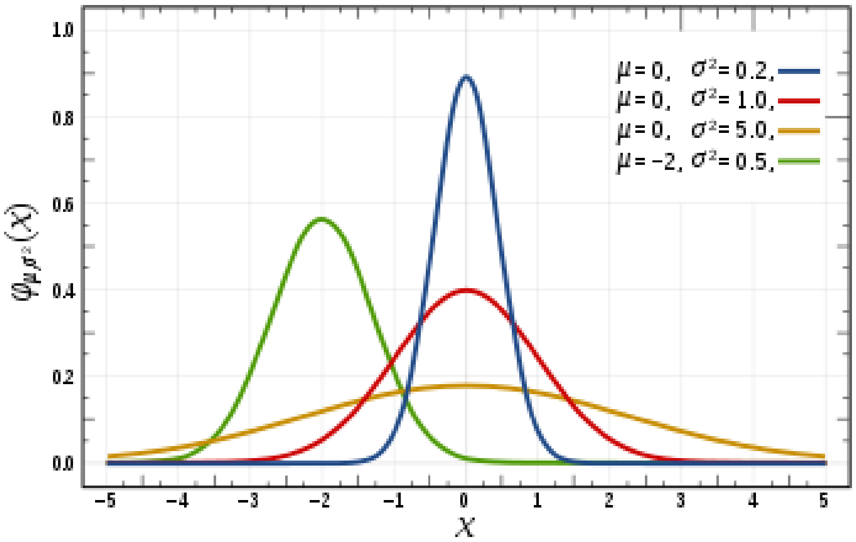
Gaussian distribution

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\text{mean} = \mu$$

$$\text{variance} = \sigma^2$$

$$\text{standard deviation} = \sigma$$



Fundamental Metabolomics

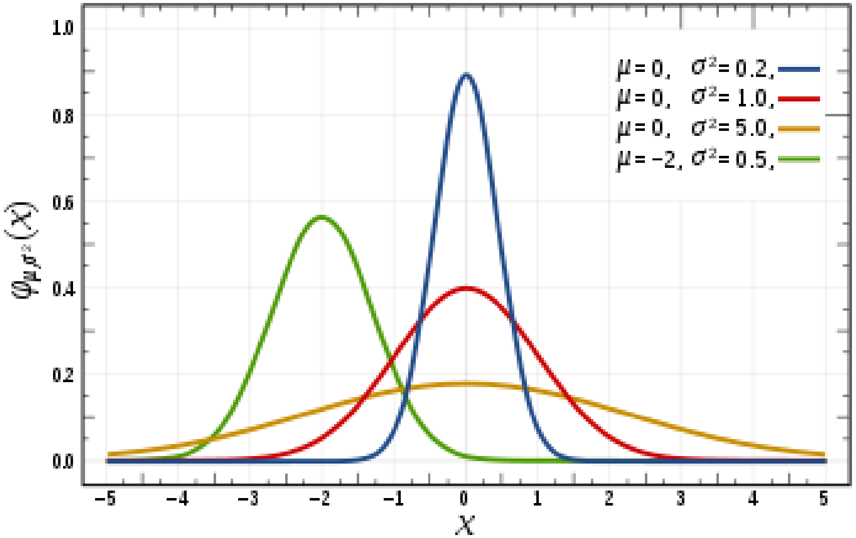
Gaussian distribution

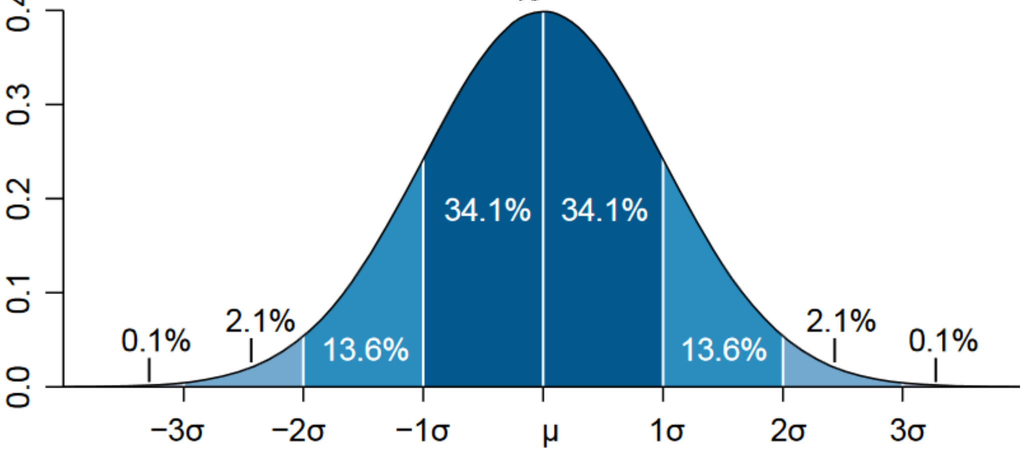
$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^{2}}$$

$$\text{mean} = \mu$$

$$\text{variance} = \sigma^{2}$$

$$\text{standard deviation} = \sigma$$





Fundamental Metabolomics

Sample mean: $\bar{\mathrm{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$

Sample standard deviation: $S = \sqrt{S^2}$

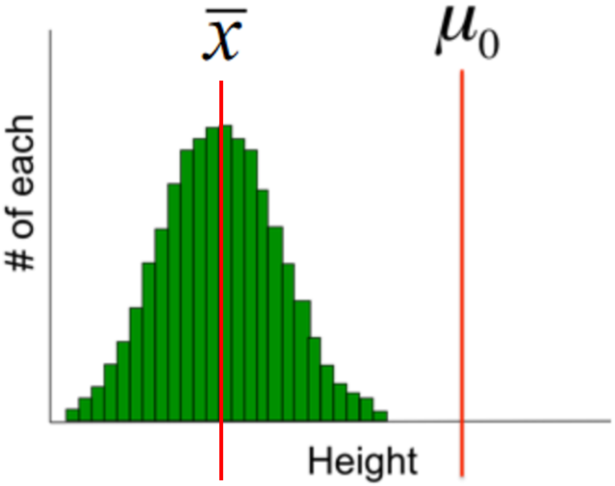
Fundamental Metabolomics

One-sample t-test: is the sample belong to a known population

Null hypothesis H_0: $\mu = \mu_0$

Test statistic: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

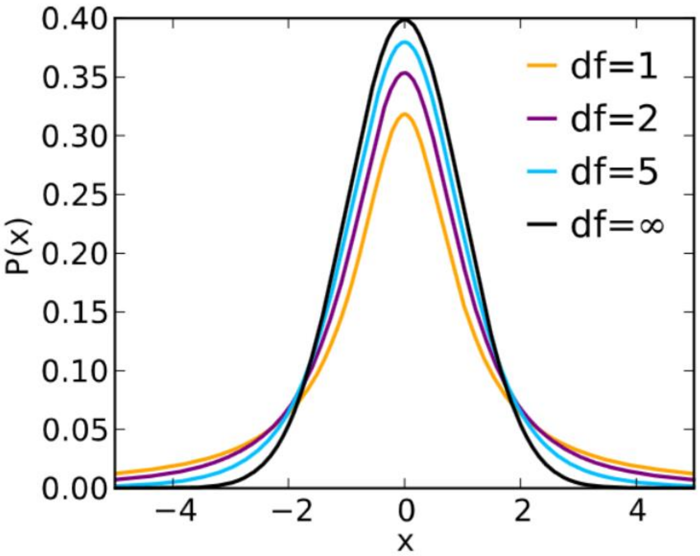
Sample standard deviation: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$



Fundamental Metabolomics

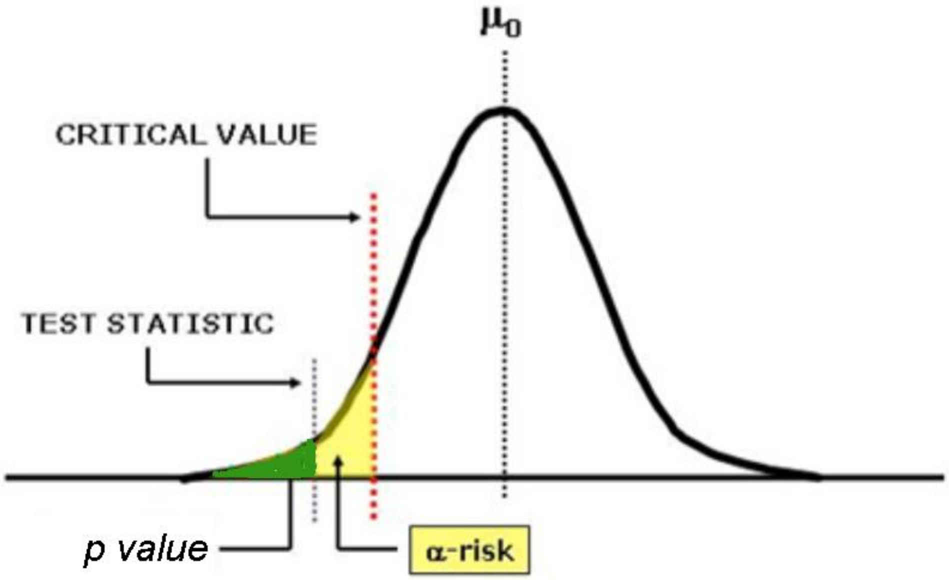
Test statistic: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

The test statistic *t* follows a student's *t* distribution.



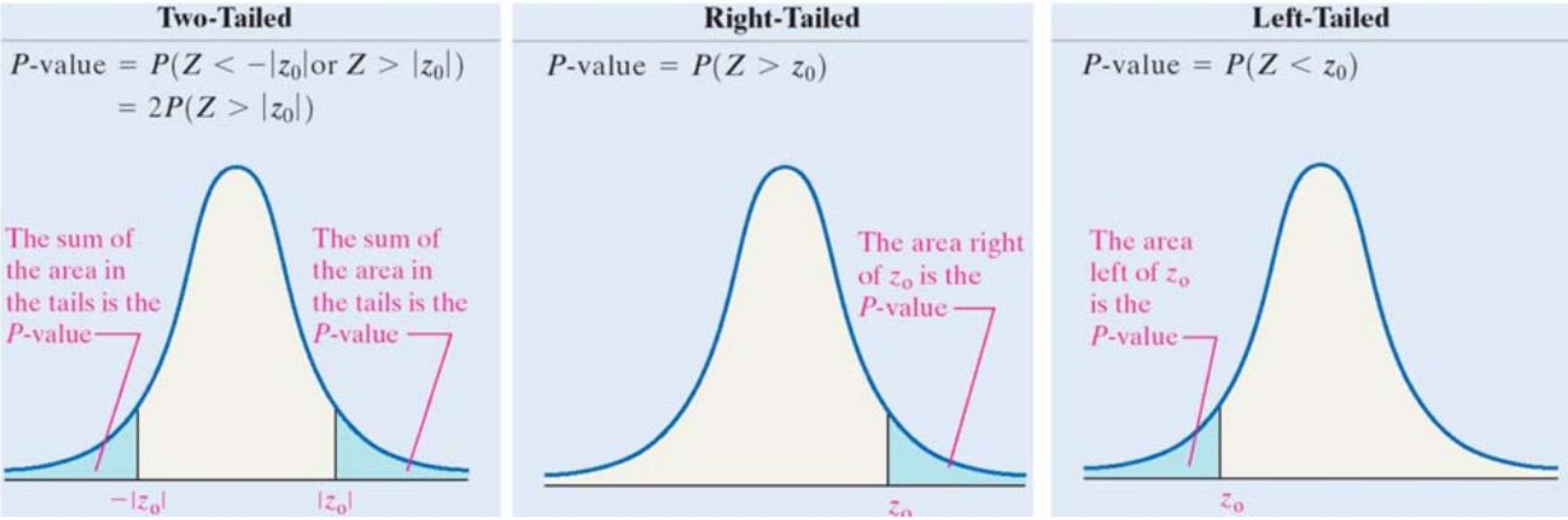
Fundamental Metabolomics

- P-value is the area in the tail of a probability distribution.
- If p-value is smaller than the critical significance level **α,** the null hypothesis is rejected and the tested sample is significant different from the population



Fundamental Metabolomics

- P-value is the area in the tail of a probability distribution.
- If p-value is smaller than the critical significance level **α,** the null hypothesis is rejected and the tested sample is significant different from the population



Two-sample t-test: are the two populations different ?
e.g., spermine concentration in normal vs. colon cancer patients

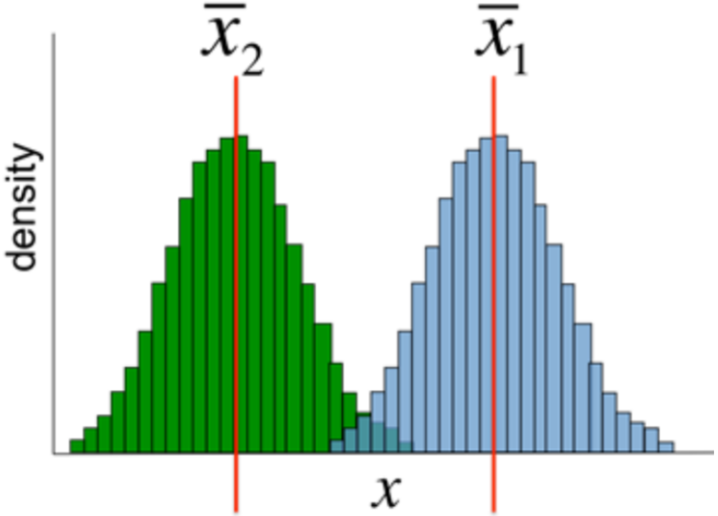
Fundamental Metabolomics

Two-sample t-test: are the two populations different ?
e.g., spermine concentration in normal vs. colon cancer patients

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$

Alternative hypethesis $H_1: \mu_1 - \mu_2 \neq 0$

Test statistic: $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$



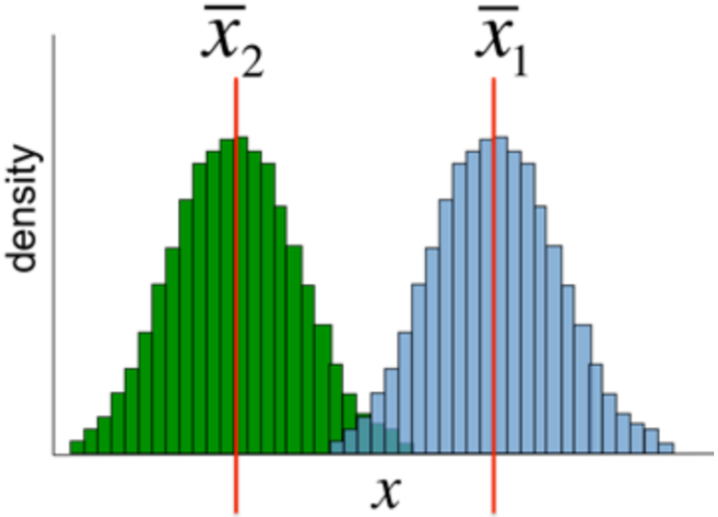
Fundamental Metabolomics

Two-sample t-test: are the two populations different ?
e.g., spermine concentration in normal vs. colon cancer patients

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$

Alternative hypethesis $H_1: \mu_1 - \mu_2 \neq 0$

$$\text{Test statistic: } t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$



Equivalent statement:

- If the t-test statistic gives a p = 0.01, it means there is only 1% of the chance that these two populations are the same, i.e. is statistically significant

Student’s t-Test

- **Paired t-Test**: Measurements made on the same individuals before and after the treatments. e.g., what is the effect of a treatment.

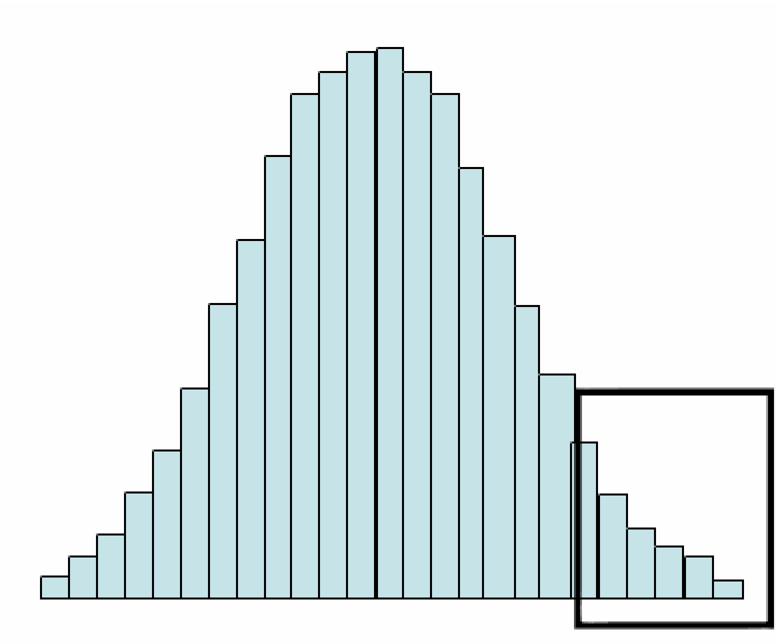
- **Unpaired t-Test**: used for the statistical analysis of 2 groups of unrelated samples

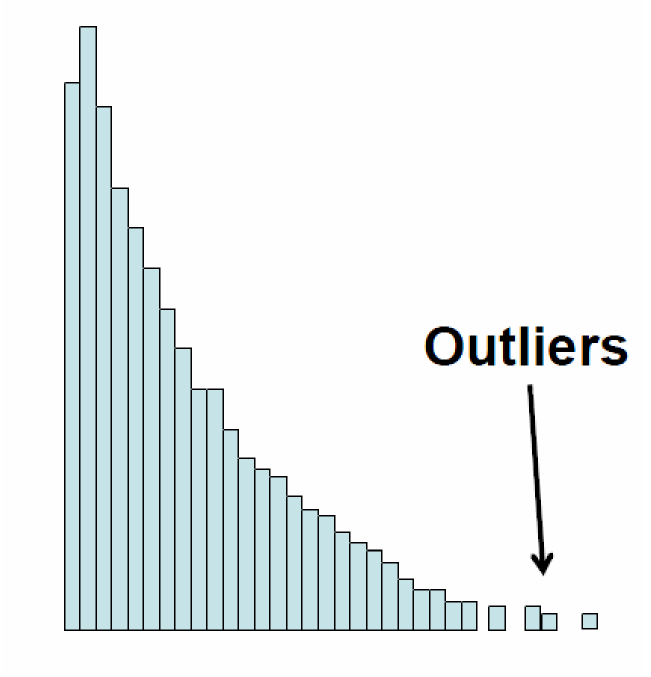
Student’s t-Test

- **Parametric t-test**: If the two population are normally distributed

- **Non-parametric t-test (Mann-Whitney t-test)**: if the two populations are non-normally distributed. More powerful and robust than the Parametric t-test. Commonly used in metabolomics

Fundamental Metabolomics


Normal distribution


Skewed distribution

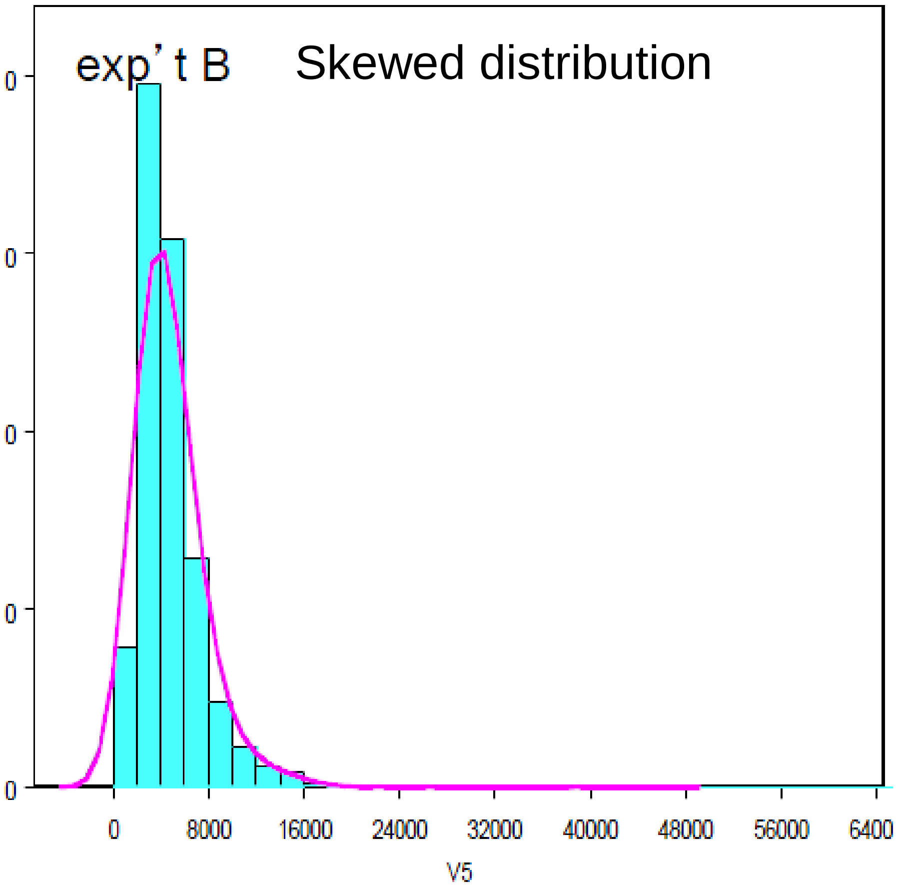
- Lots of extreme values far away from mean
- Hard to do useful statistical tests

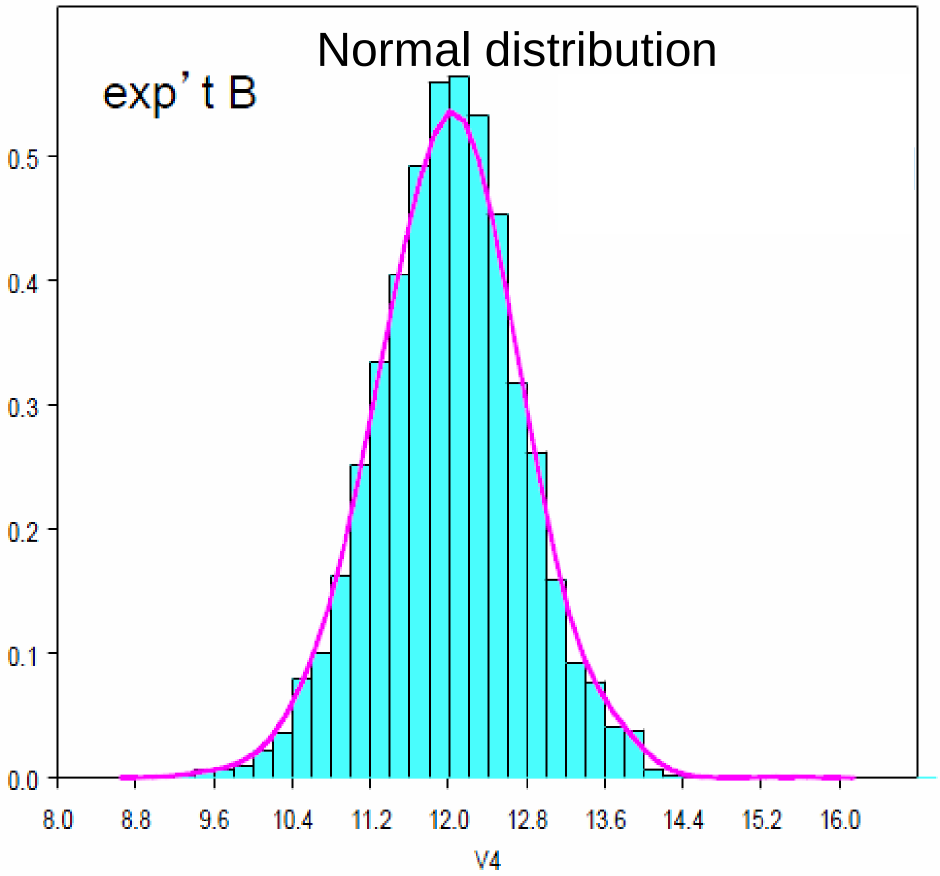
Fundamental Metabolomics

Apply a log transformation to fixing a skewed distribution

Fundamental Metabolomics

Apply a log transformation to fixing a skewed distribution





Fundamental Metabolomics

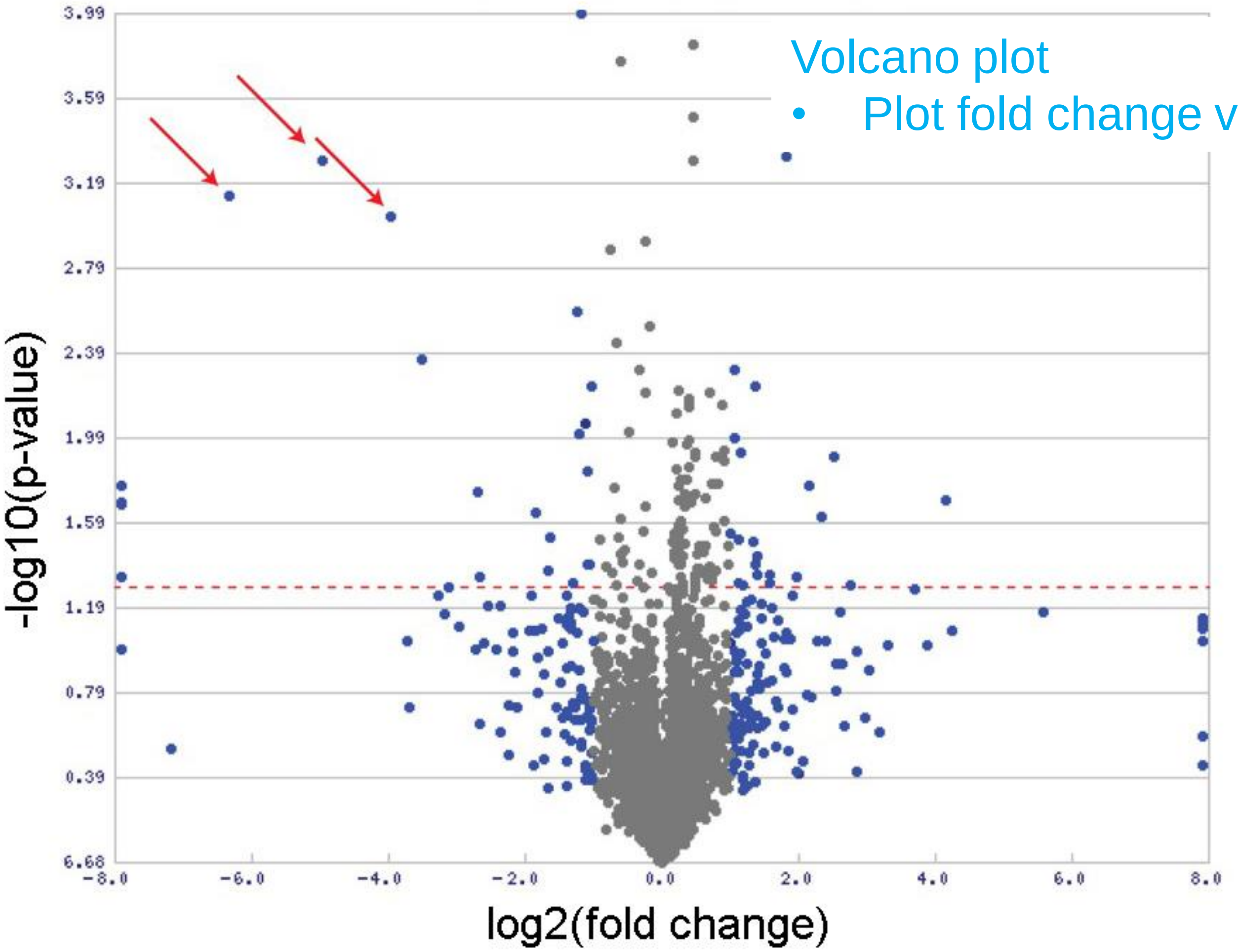
Analysis of variance (ANOVA)

Fundamental Metabolomics

Analysis of variance (ANOVA)

- Used to determine if 3 or more populations are different
- A generalization of the t-test
- It provides a statistical test of whether or not the means of several groups are all equal or whether 3+ groups of values are different

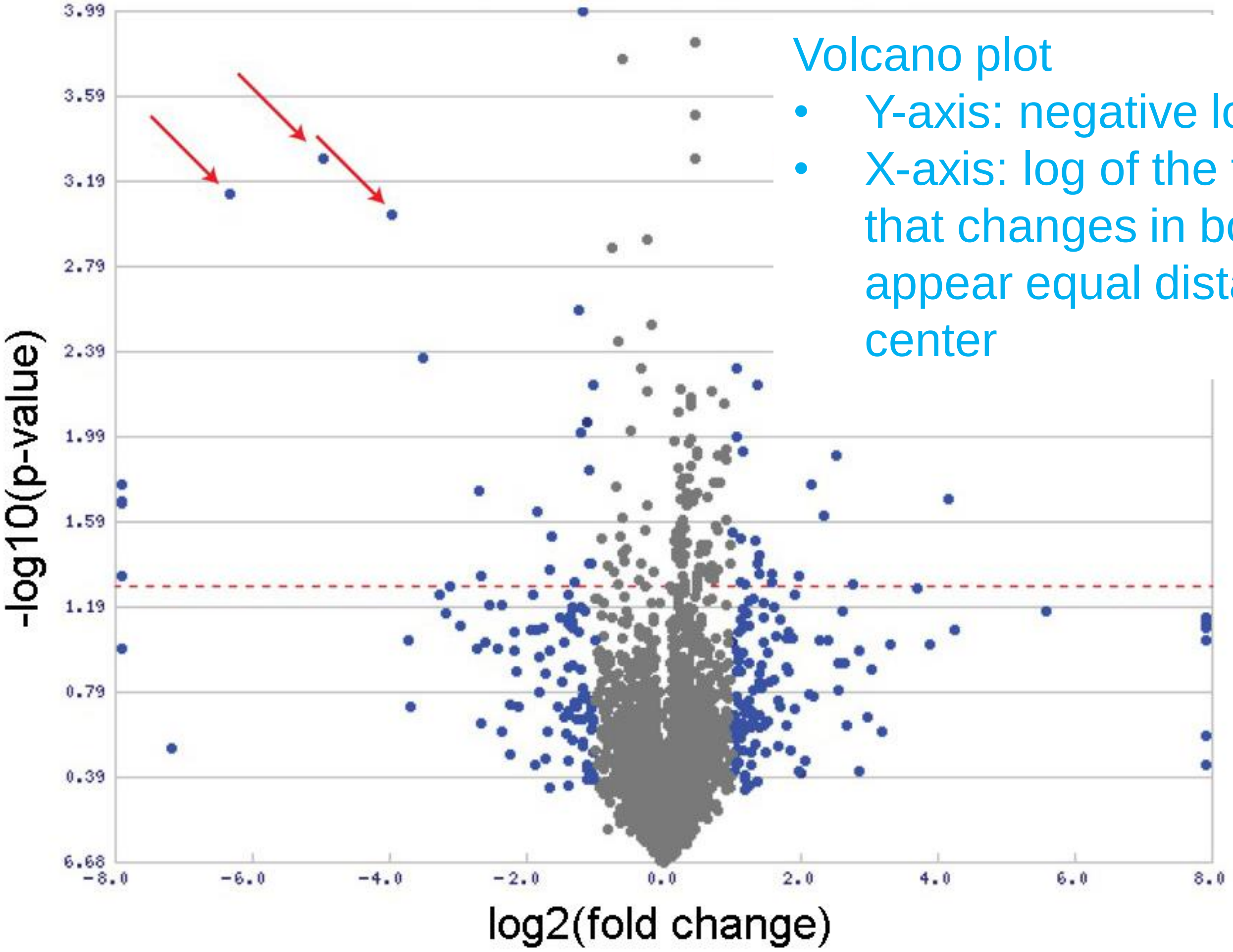
Fundamental Metabolomics



Volcano plot

- Plot fold change vs. significance

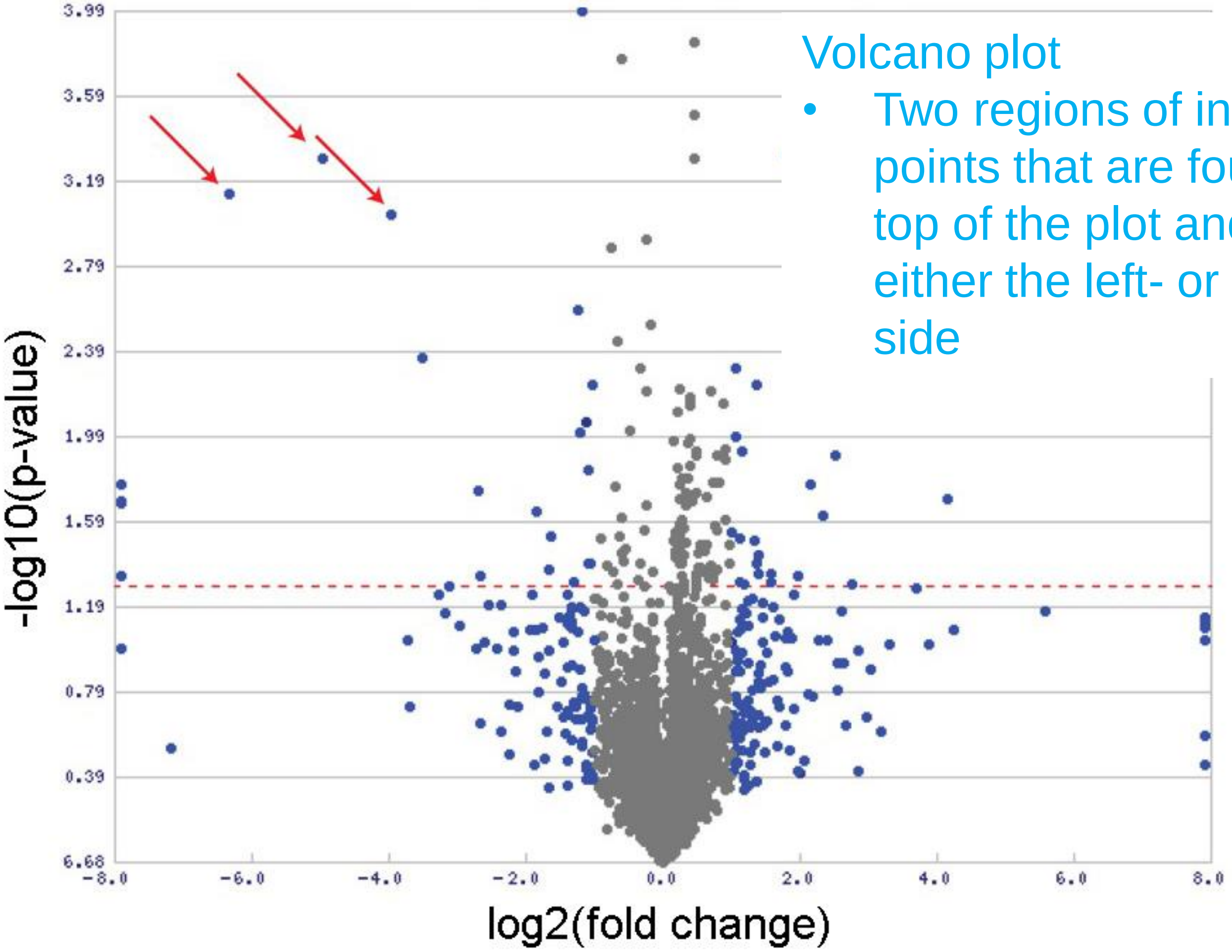
Fundamental Metabolomics



Volcano plot

- Y-axis: negative log of the p-value
- X-axis: log of the fold change so that changes in both directions appear equal distance from the center

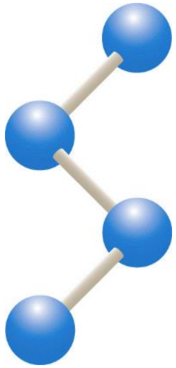
Fundamental Metabolomics



Volcano plot

- Two regions of interest: those points that are found towards the top of the plot and are far to either the left- or the right- hand side

May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Statistical Analysis***

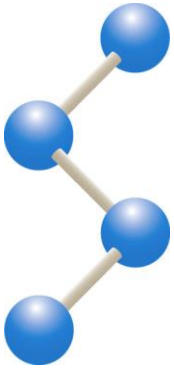
May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

May 16th

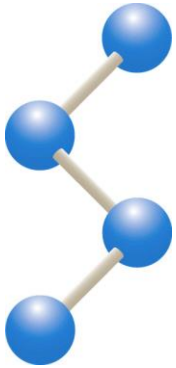


Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

May 16th
---- 12:30 pm Begin ----
---- 3:00 pm Break ----
---- 04:30 pm Finish ----

May 16th



Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----

Univariate vs. Multivariate Analysis

- **Univariate:** analysis of each variable separately (e.g., unpaired/paired t-test)

- **Multivariate: simultaneous analysis of N variables**

Univariate vs. Multivariate Analysis

- **Univariate:** analysis of each variable separately (e.g., unpaired/paired t-test)

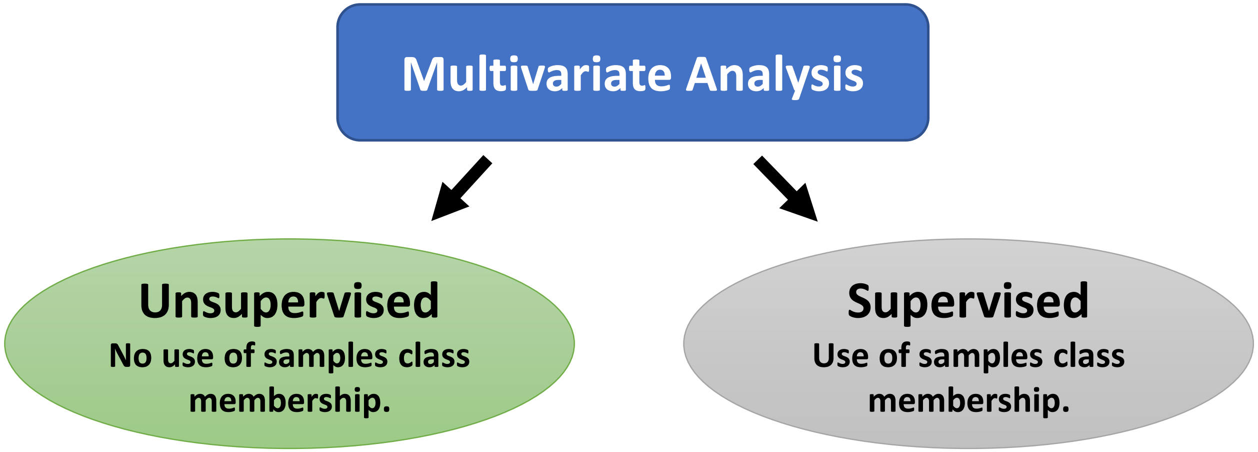
- **Multivariate: simultaneous analysis of N variables**

Multivariate Analysis

Univariate vs. Multivariate Analysis

- **Univariate:** analysis of each variable separately (e.g., unpaired/paired t-test)

- **Multivariate: simultaneous analysis of N variables**



Multivariate Analysis

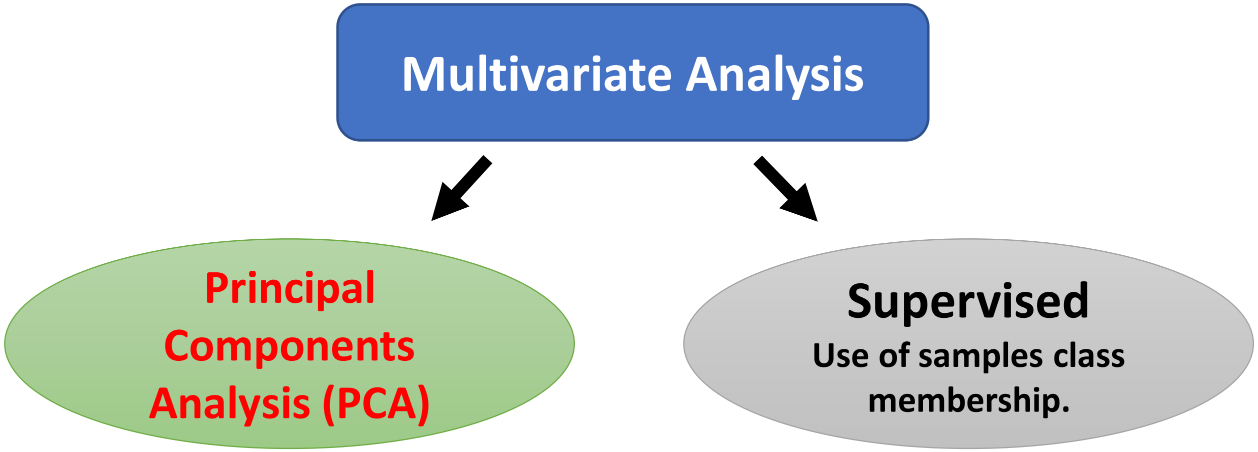
Unsupervised
No use of samples class membership.

Supervised
Use of samples class membership.

Univariate vs. Multivariate Analysis

- **Univariate:** analysis of each variable separately (e.g., unpaired/paired t-test)

- **Multivariate: simultaneous analysis of N variables**



Multivariate Analysis

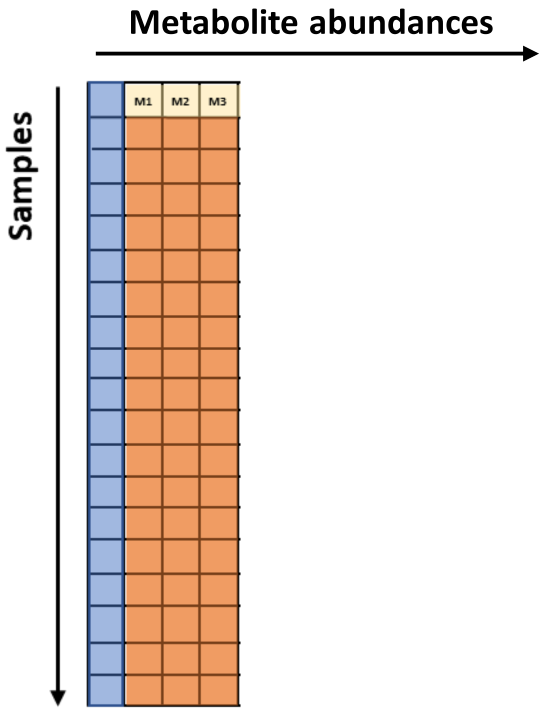
Principal Components Analysis (PCA)

Supervised
Use of samples class membership.

Visualization of Data Variability

Targeted Metabolomics

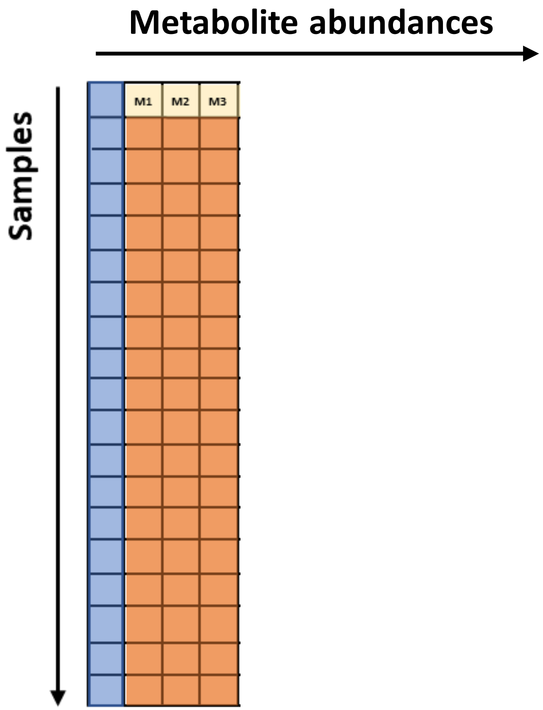
- E.g. 3 metabolite concentrations measured for each sample.



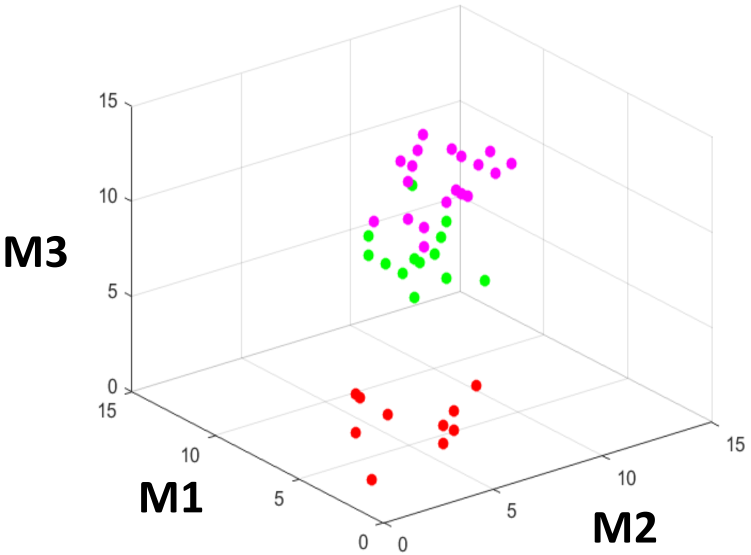
Visualization of Data Variability

Targeted Metabolomics

- E.g. 3 metabolite concentrations measured for each sample.



3-D Plot

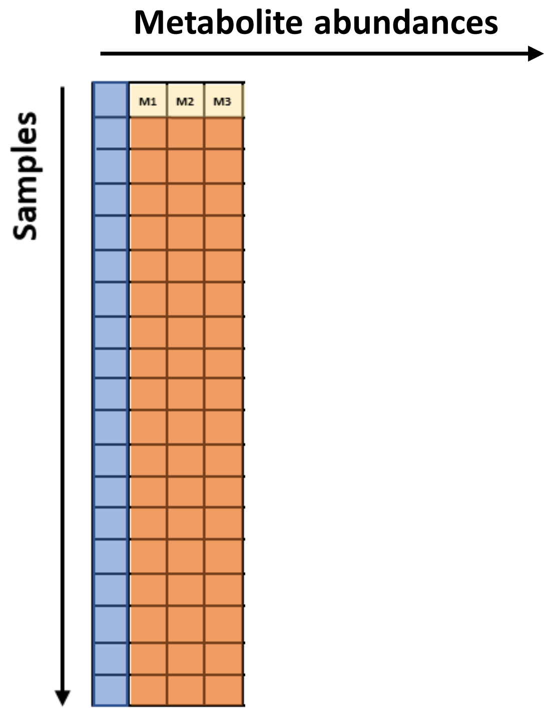


For 3 variables, we can estimate and visualize the distribution of the variability within our data.

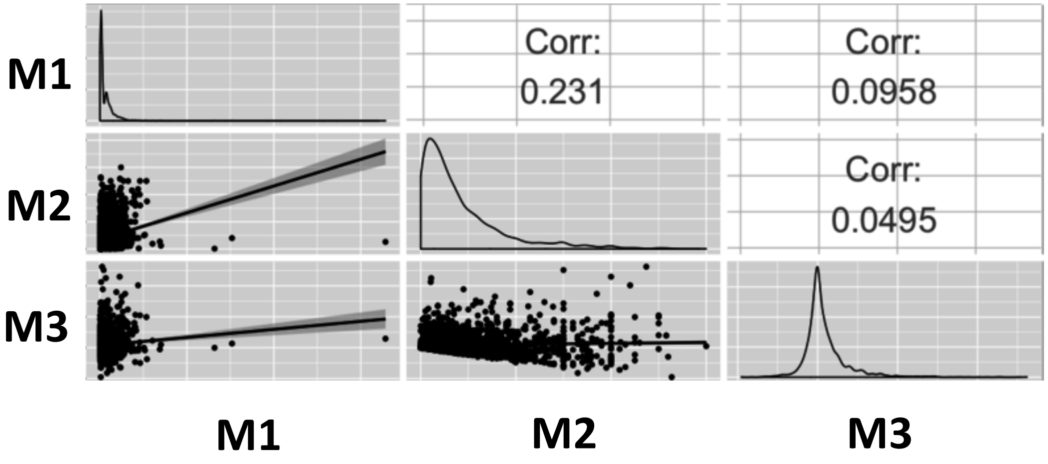
Visualization of Data Variability

Targeted Metabolomics

- E.g. 3 metabolite concentrations measured for each sample.



2-D Correlation Matrix

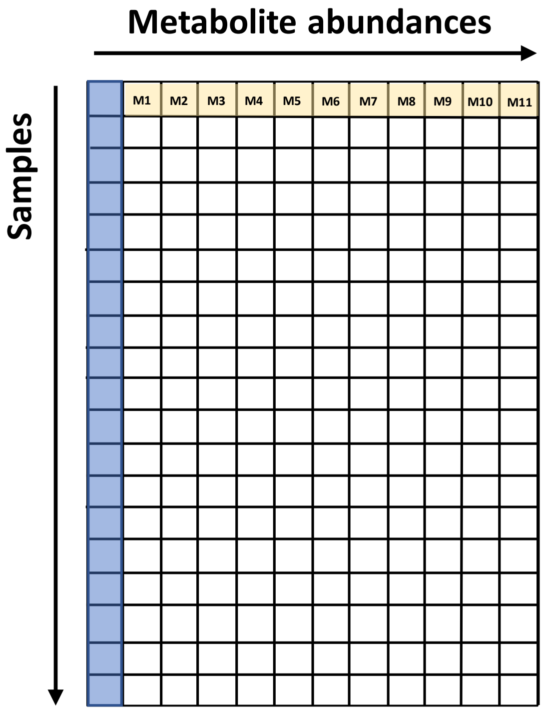


For 3 variables, we can estimate and visualize the distribution of the variability within our data by plotting pairs of variables.

Data Dimensionality in Metabolomics

Targeted Metabolomics

- **Generally < 100 metabolite concentrations measured for each sample.**



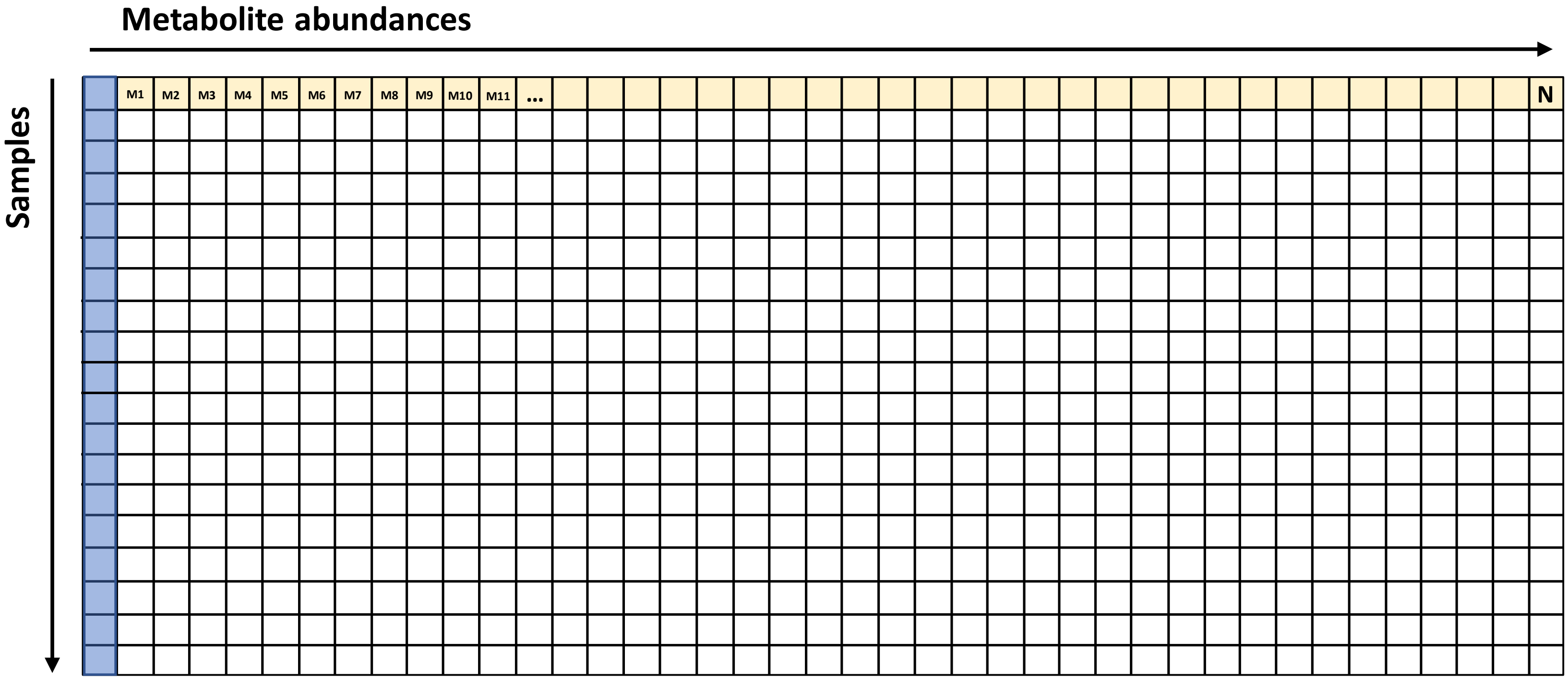
Data Dimensionality in Metabolomics

Targeted Metabolomics

- Generally < 100 metabolite concentrations measured for each sample.

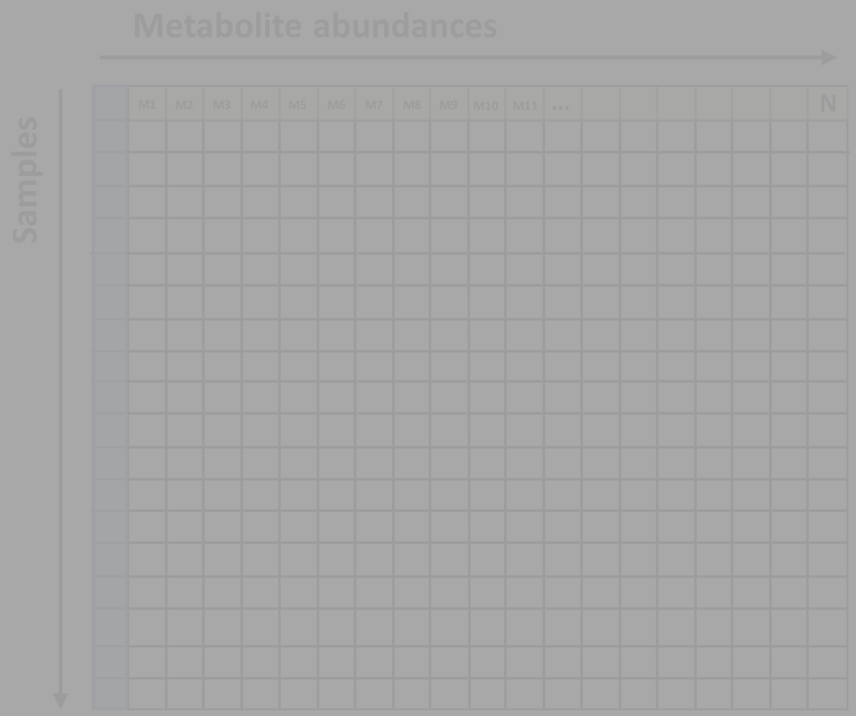
Untargeted Metabolomics

- N >> 100 metabolic feature abundances measured for each sample.



The Data Visualization Problem in Metabolomics

- **Metabolomics provides high-dimensional data sets (big data).**



The Data Visualization Problem in Metabolomics

- Metabolomics provides high-dimensional data sets (big data).

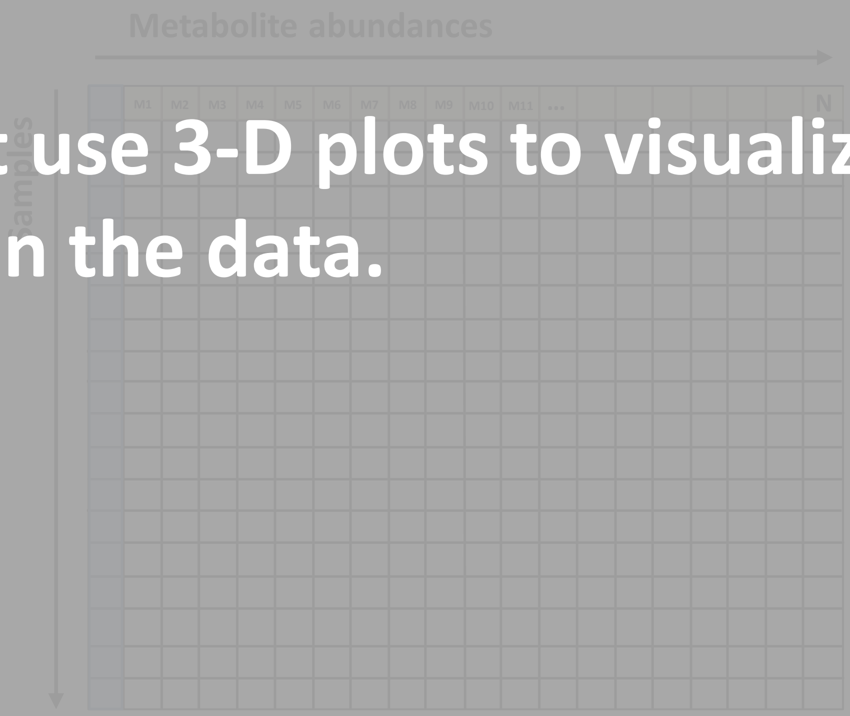
- We can not use 3-D plots to visualize the variability in the data.

Targeted Metabolomics

- Generally [illegible] metabolite concentrations measured for each sample

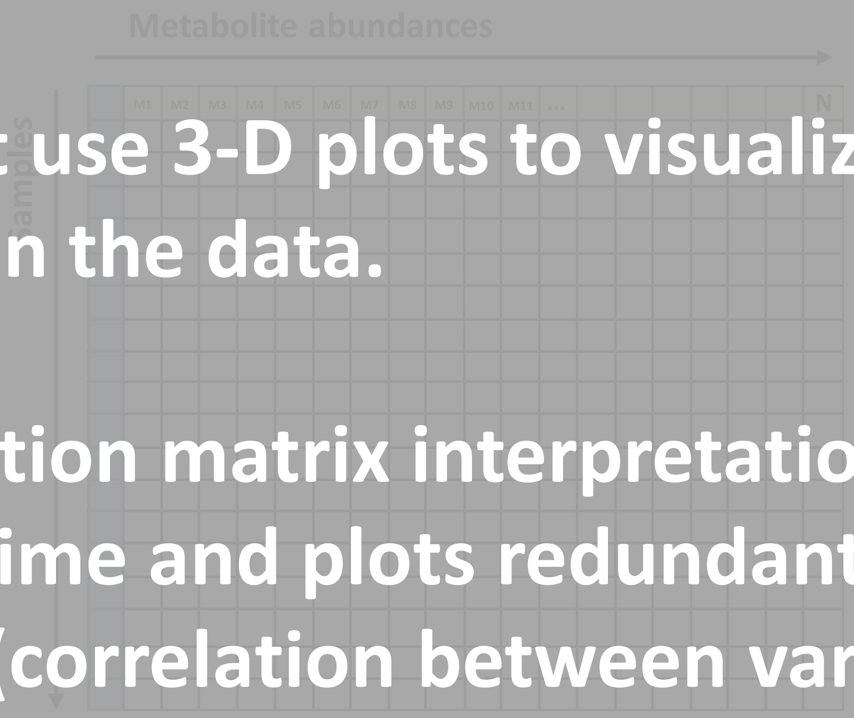
Untargeted Metabolomics

- Usually [illegible] metabolite feature abundances measured for each sample



The Data Visualization Problem in Metabolomics

- Metabolomics provides high-dimensional data sets (big data).
- We can not use 3-D plots to visualize the variability in the data.
- 2-D correlation matrix interpretation requires too much time and plots redundant variability (correlation between variables).



Why PCA of Metabolomic Data?

1. Visual estimation of the latent variability in the data set at a glance (e.g. **spontaneous group clustering**).

Why PCA of Metabolomic Data?

1. Visual estimation of the latent variability in the data set at a glance (e.g. **spontaneous group clustering**).
2. **Outliers detection** (e.g. quality control).

Why PCA of Metabolomic Data?

1. Visual estimation of the latent variability in the data set at a glance (e.g. **spontaneous group clustering**).
2. **Outliers detection** (e.g. quality control).
3. **Interpretation of the variability** in high-dimensional data by using all the variables simultaneously.

Before PCA: Data Scaling

PCA is a maximum variance projection method, therefore variables with larger range can bias the model.

SCALING and MEAN CENTERING before PCA

e.g. Auto-Scaling:

- Mean Centering: the average value for each variable is calculated and subtracted from the data.

Before PCA: Data Scaling

PCA is a maximum variance projection method, therefore variables with larger range can bias the model.

SCALING and MEAN CENTERING before PCA

e.g. Auto-Scaling:

- Mean Centering: the average value for each variable is calculated and subtracted from the data.
- Each metabolite measurement is multiplied for its **1/SD**. SD: standard deviation, consequently each variable has equal (unit) variance.

Before PCA: Data Scaling

PCA is a maximum variance projection method, therefore variables with larger range can bias the model.

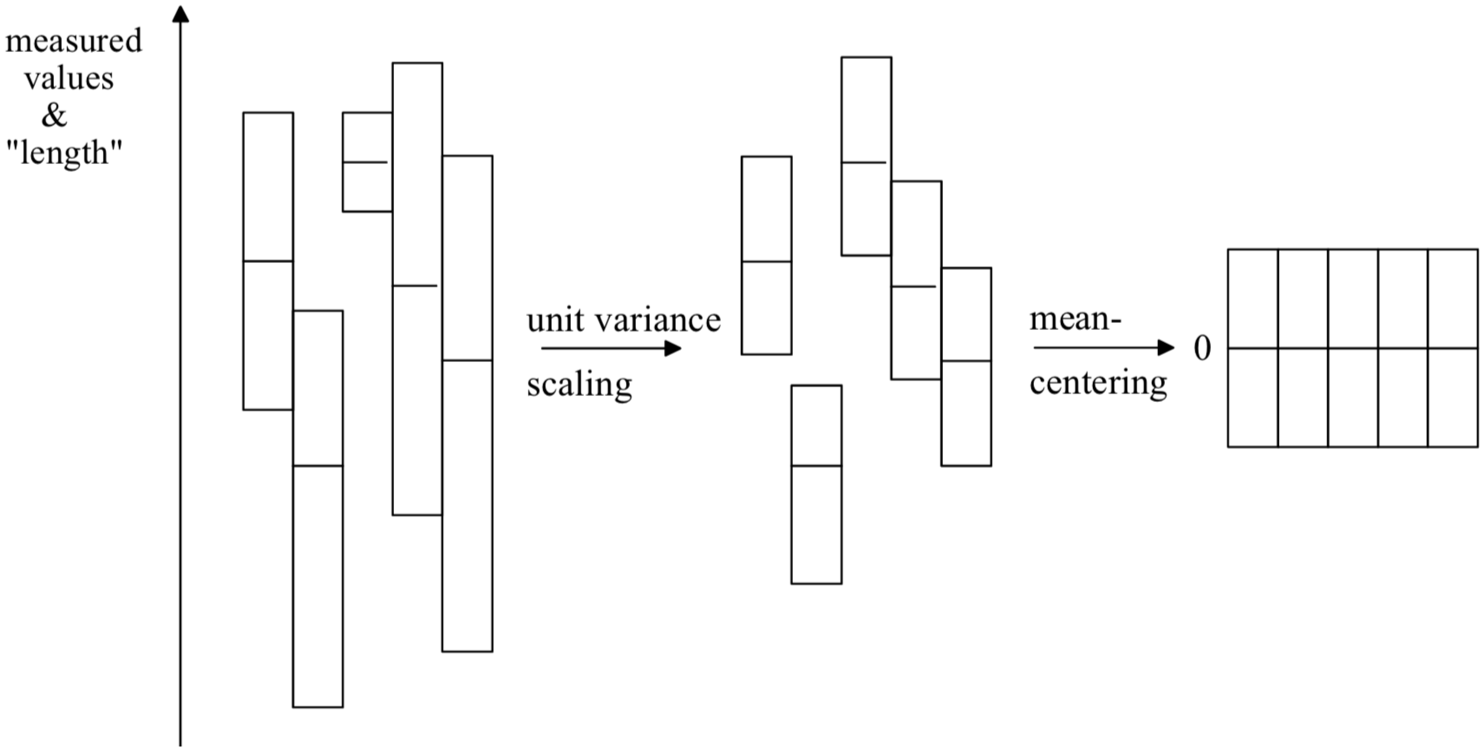
SCALING and MEAN CENTERING before PCA

e.g. Auto-Scaling:

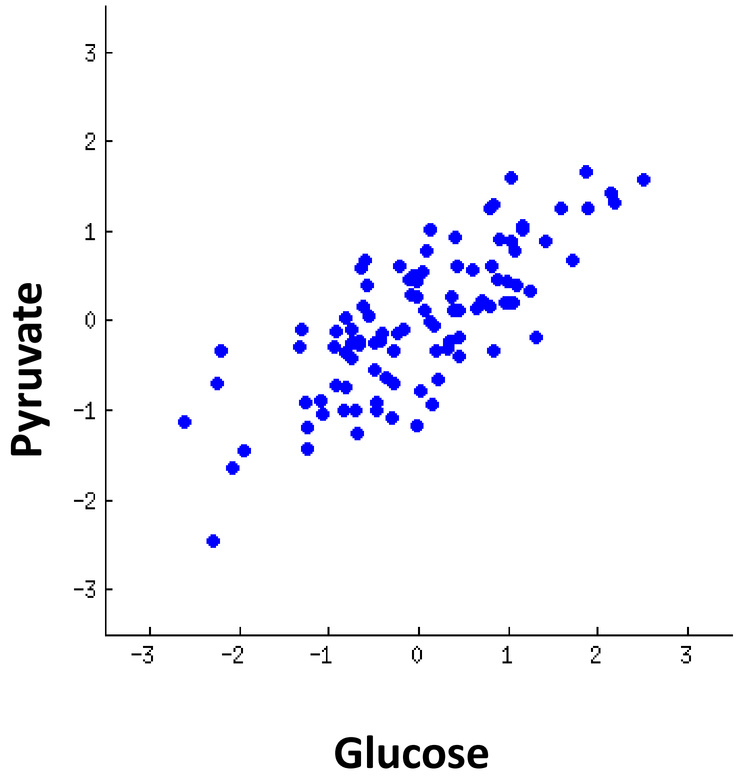
- Mean Centering: the average value for each variable is calculated and subtracted from the data.
- Each metabolite measurement is multiplied for its **1/SD**. SD: standard deviation, consequently each variable has equal (unit) variance.

Other options: Pareto scaling, Log Transformation

Data Auto-Scaling

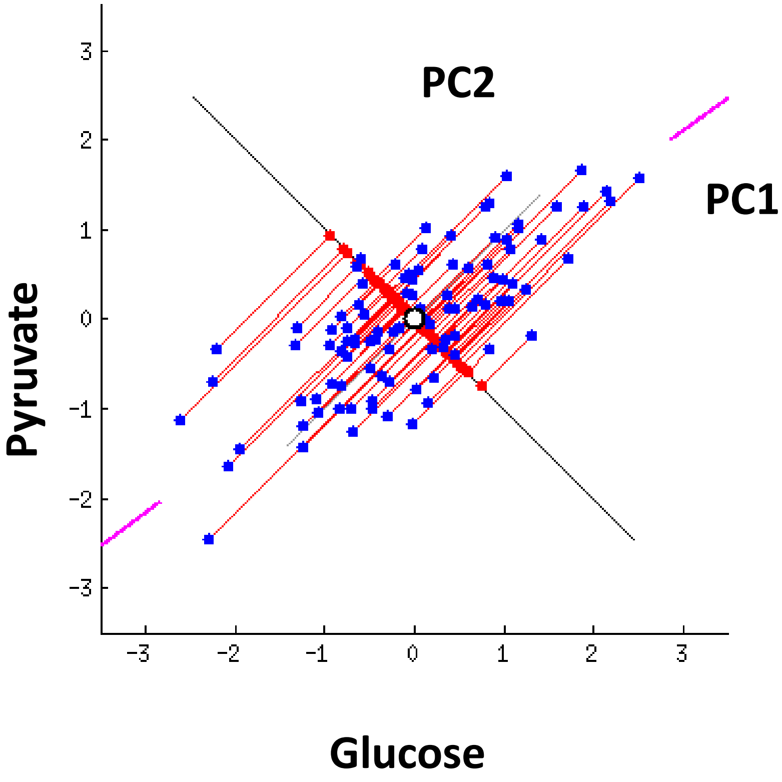


How Does PCA Work?



- **PCA** performs an **orthogonal transformation** to convert a set of observations of **linearly** correlated variables in a **set of linearly uncorrelated variables (Principal Components, or eigenvectors)**

How Does PCA Work?

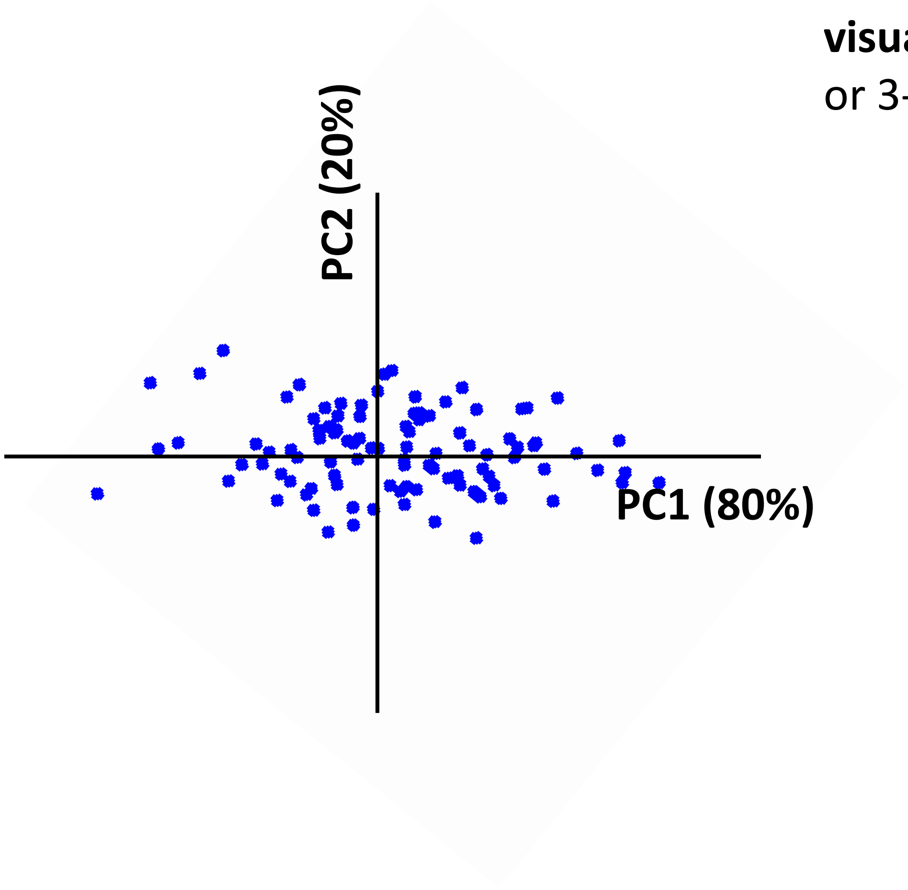


- **PCA** performs an **orthogonal transformation** to convert a set of observations of **linearly** correlated variables in a **set of linearly uncorrelated variables (Principal Components, or eigenvectors)**

- **It computes new coordinates (eigenvectors or Principal Components) from the original coordinates.**

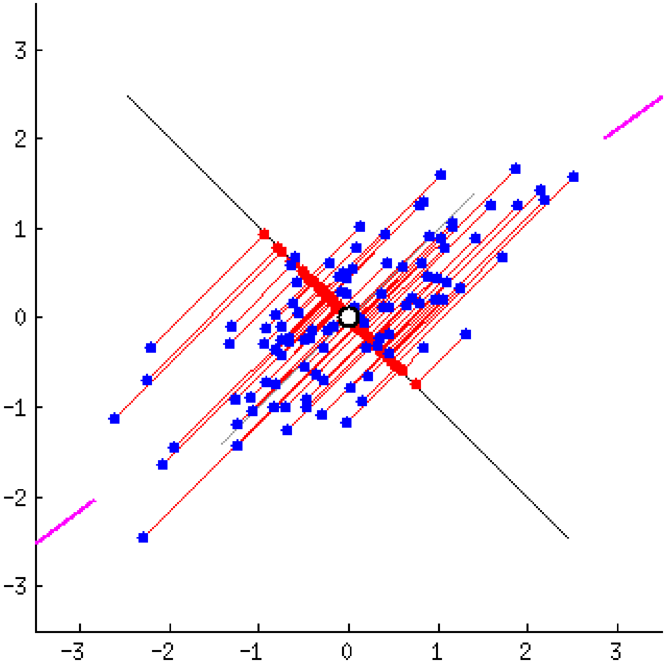
- **The PCs are ordered according to the amount of explained variability.**

How Does PCA Work?



- We can **discard low-variance variables to visualize the variability** in our data set in a 2-D or 3-D plot.

PCA of High-Dimensional Metabolomics Data



- For high-dimensional data sets, this operation **is repeated for all the dimensions in our experiment.**

PCA Scree Plot

- **PCA** provides a **list of Principal Components** with the related **amount of explained variability** in the data set (eigenvalues).

PCA Scree Plot

- **PCA** provides a **list of Principal Components** with the related **amount of explained variability** in the data set (eigenvalues).

- **Scree plot: visual aid to determine the number of PCs to plot.**

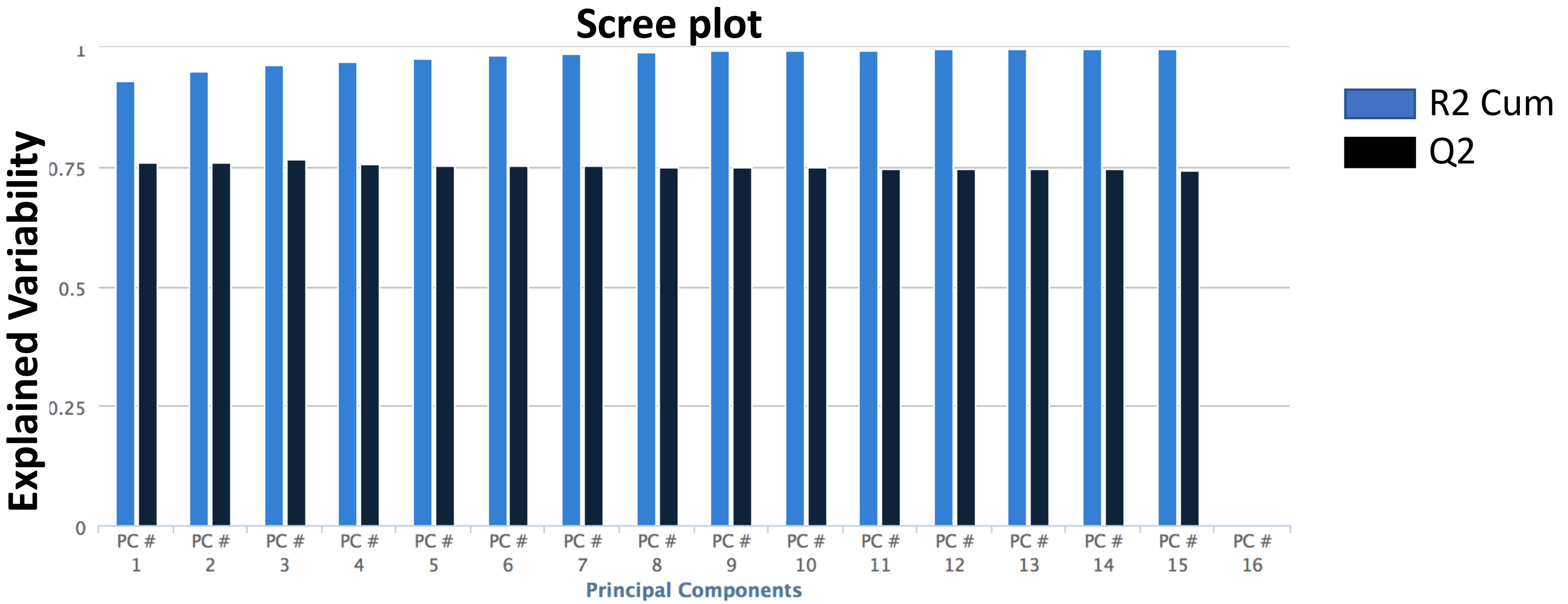
Example 1: Colon Cancer Study

- **PCA** provides a **list of Principal Components** with the related **amount of explained variability** in the data set (eigenvalues).

- **Scree plot: visual aid to determine the number of PCs to plot.**

e.g. Colon Cancer Vs. Normal Colon Tissue in Human.

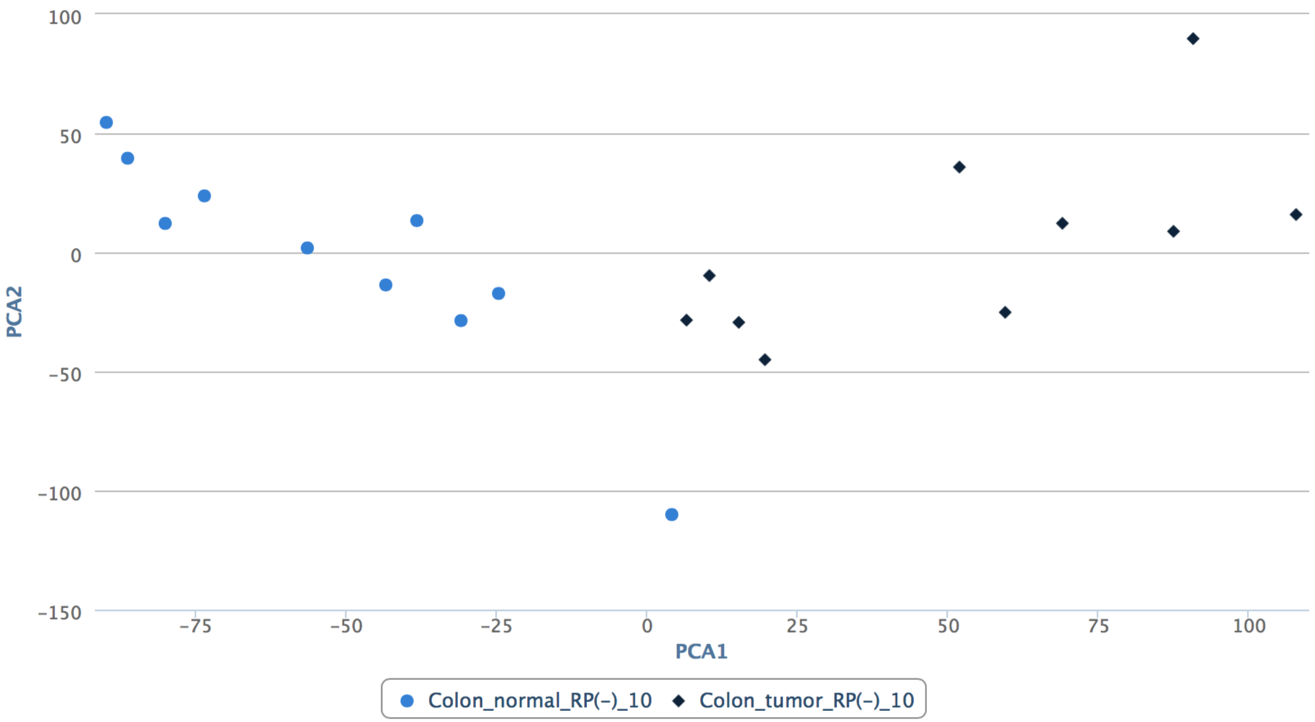
17130 total aligned features, 10 samples/class



PCA Score Plot

- Each observation can be projected on the PCs in order to get a coordinate value along the PC-line: this value is known as a score. Scores for PCs can be plotted to visualize variability.

e.g. score plot of the first 2 PCs define a model plane.

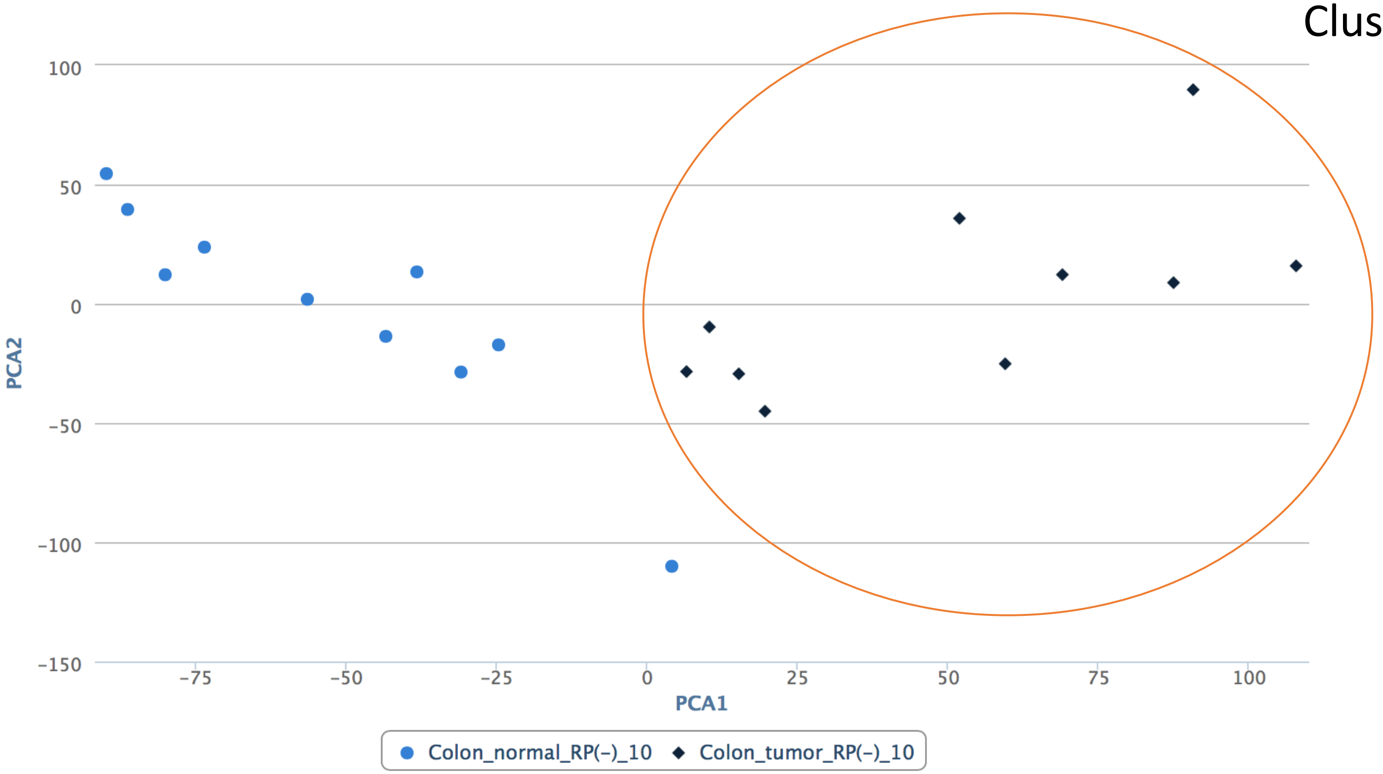


PCA Score Plot

- **Each observation can be projected on the PCs in order to get a coordinate value along the PC-line: this value is known as a score. Scores for PCs can be plotted to visualize variability.**

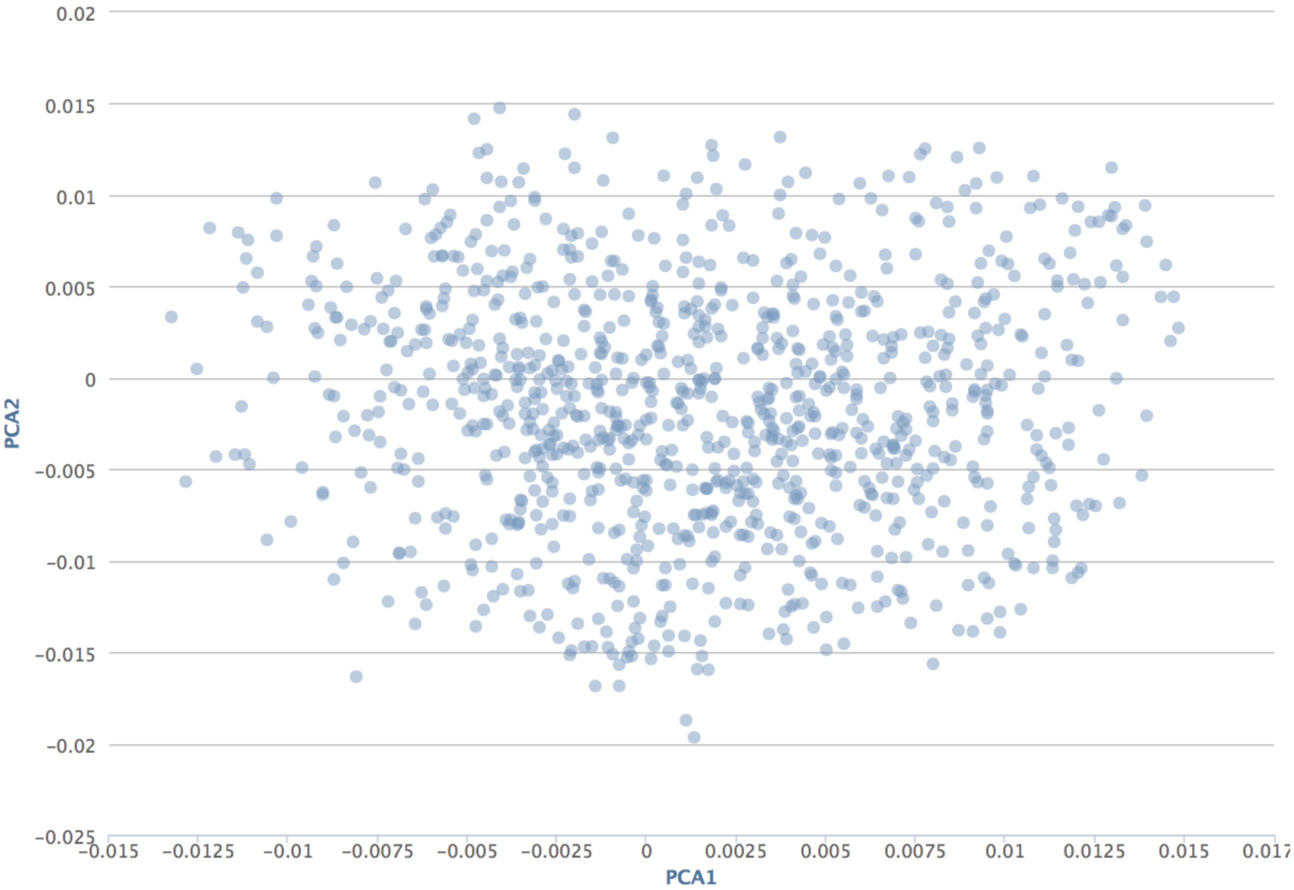
e.g. score plot of the first 2 PCs define a model plane.

Unsupervised Clustering



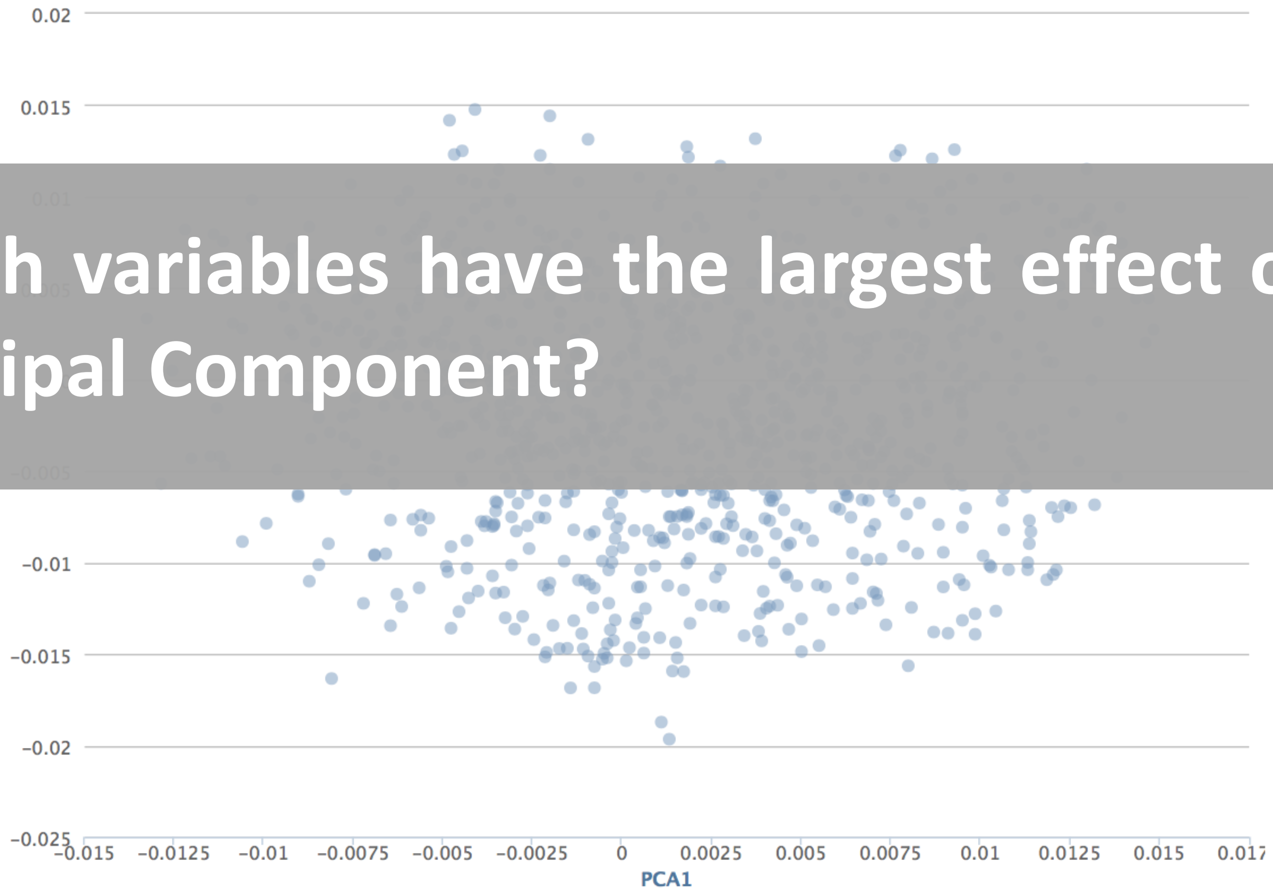
PCA Loading Plot

- The loading plot represents **the relationship** between **the original variables** and **the computed dimensions (PCs).**



PCA Loading Plot

- The loading plot represents **the relationship** between **the original variables** and **the computed dimensions (PCs).**

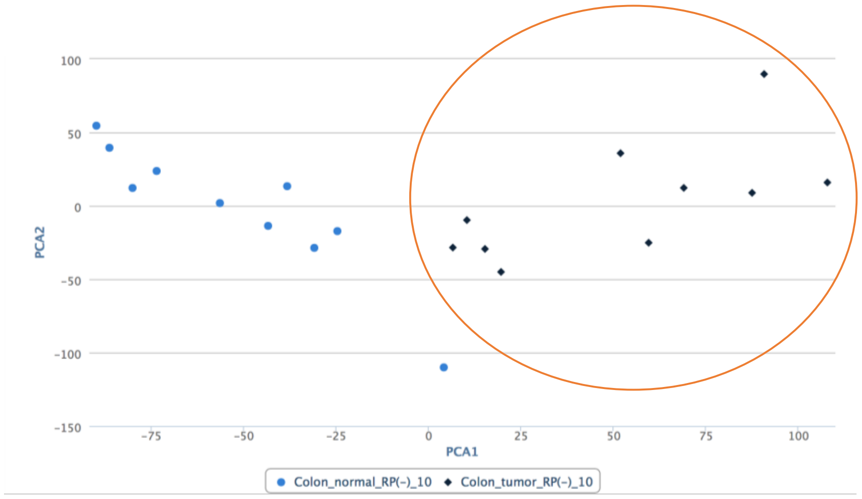


Which variables have the largest effect on each Principal Component?

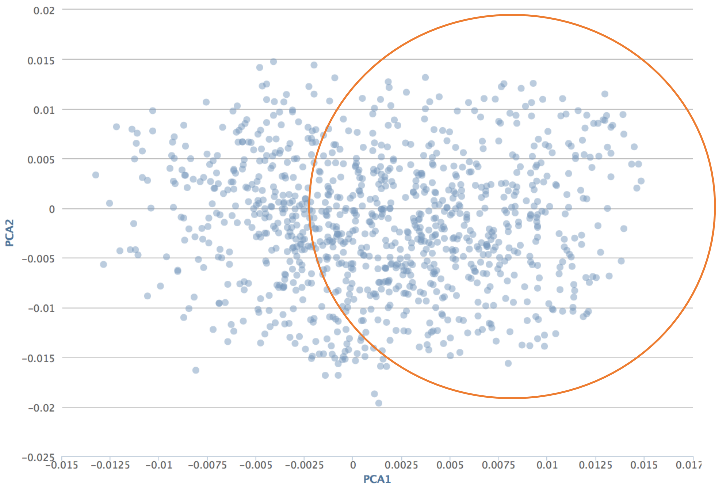
Score Vs. Loading Plot

- **Loadings range from -1 to +1. As closer to ± 1, as more influence the variable has on the component.**

Score



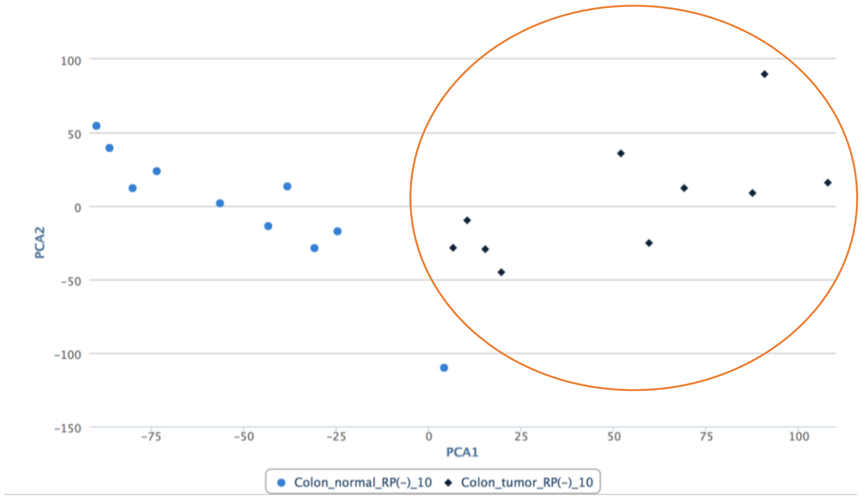
Loading



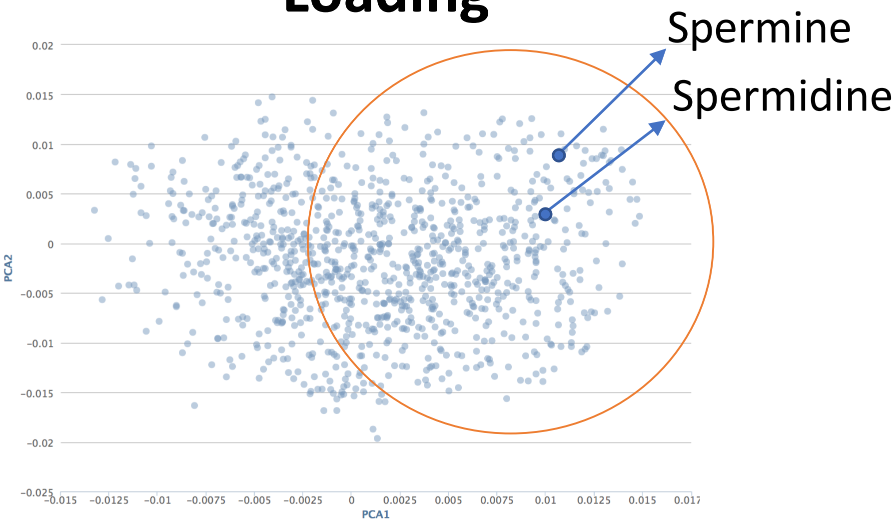
Score Vs. Loading Plot

- Loading ranges from -1 to +1. As closest to ± 1, as more influence has the variable on the component.

Score



Loading



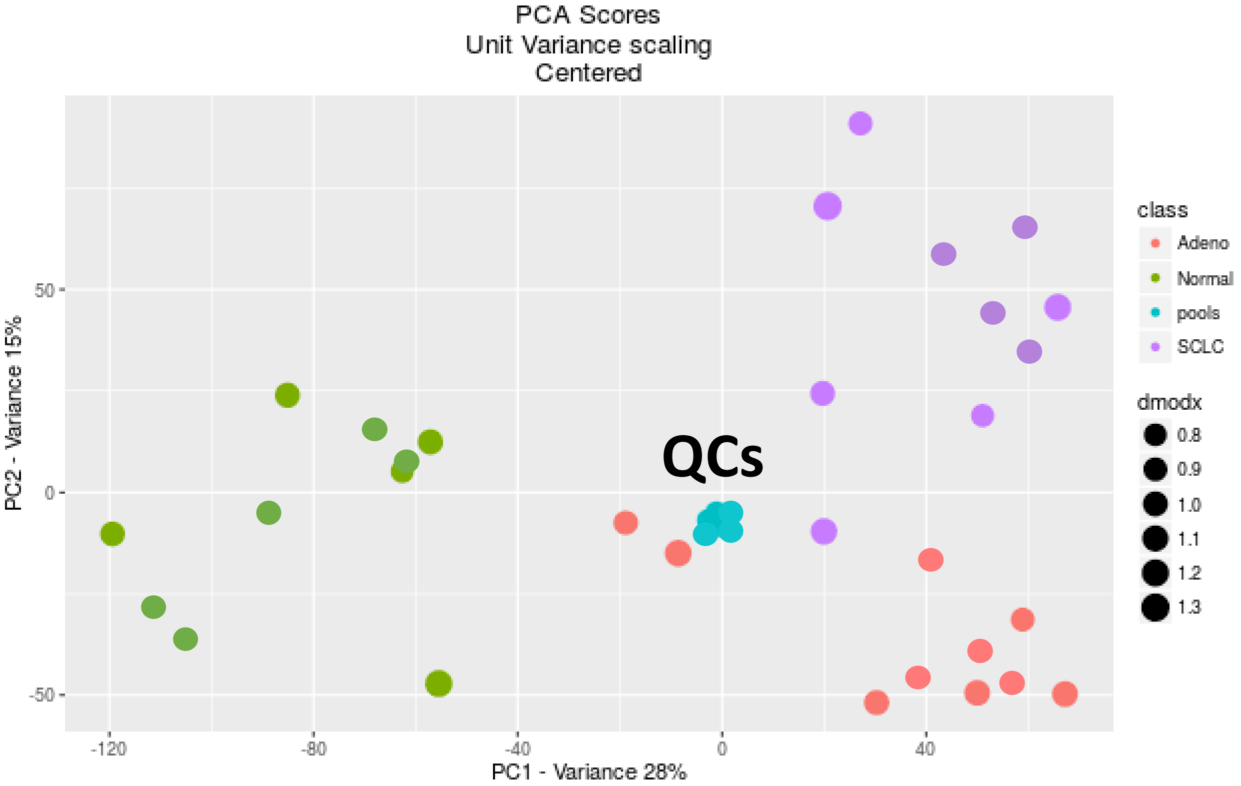
- e.g. Spermine and spermidine may be dysregulated in colon cancer.

Example 2: Quality Control

- **PCA can highlight the presence of outliers.**

This suggests **lack of reproducibility/errors in sample preparation or data acquisition. Careful interpretation of data acquired in proximity of the outlier** should be performed and the outliers excluded from further analysis.

e.g. Multigroup study: 12913 detected features, 10/Lung Adenocarcinoma, 10/Small Cells Lung Cancer, 10/Control , 10/QCs(pools).



PCA Scores
Unit Variance scaling
Centered

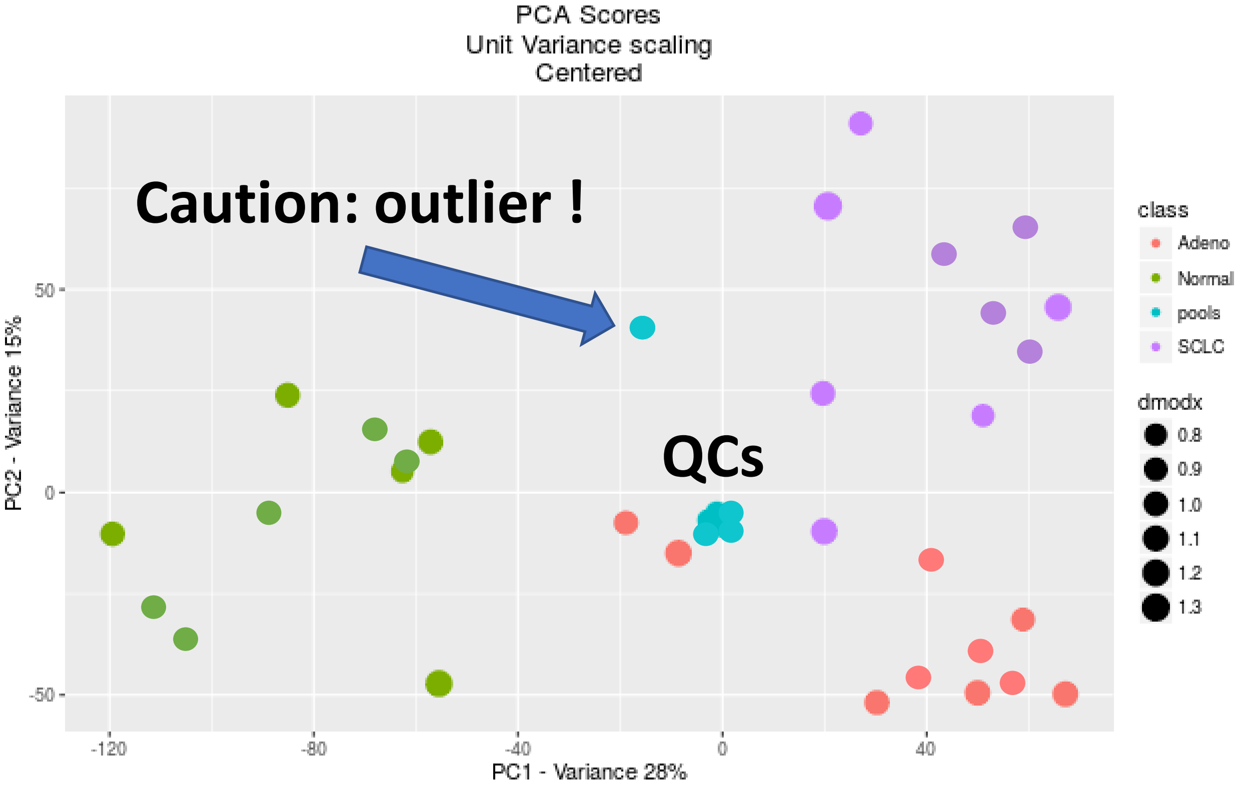
QCs

Example 2: Quality Control

- **PCA can highlight the presence of outliers.**

This suggests **lack of reproducibility/errors in sample preparation or data acquisition. Careful interpretation of data acquired in proximity of the outlier** should be performed and the outliers excluded from further analysis.

e.g. Multigroup study: 12913 detected features, 10/Lung Adenocarcinoma, 10/Small Cells Lung Cancer, 10/Control , 10/QCs(pools).



Caution: outlier !

QCs

PCA: Take Home Message

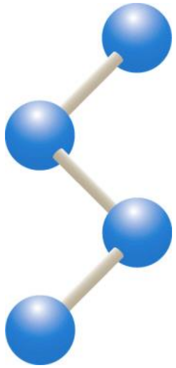
PCA is a **data dimensionality reduction** method.

Use PCA for:

1. Unravel spontaneous group **clustering /trends** in the data (score plot).

2. Evaluate **variables contribution in clustering** and variables correlation (score and loading plots).

3. Unravel strong **outliers** (score plot).



May 16th

Fundamental Metabolomics

- ***Objectives and Challenges***
- ***Experimental Design***
- ***Sample Preparation and Chromatography***
- ***Untargeted Metabolomics***
- ***Targeted Metabolomics***
- ***Metabolite Databases and Informatics***
- ***Statistical Analysis***

May 16th

---- 12:30 pm Begin ----

---- 3:00 pm Break ----

---- 04:30 pm Finish ----