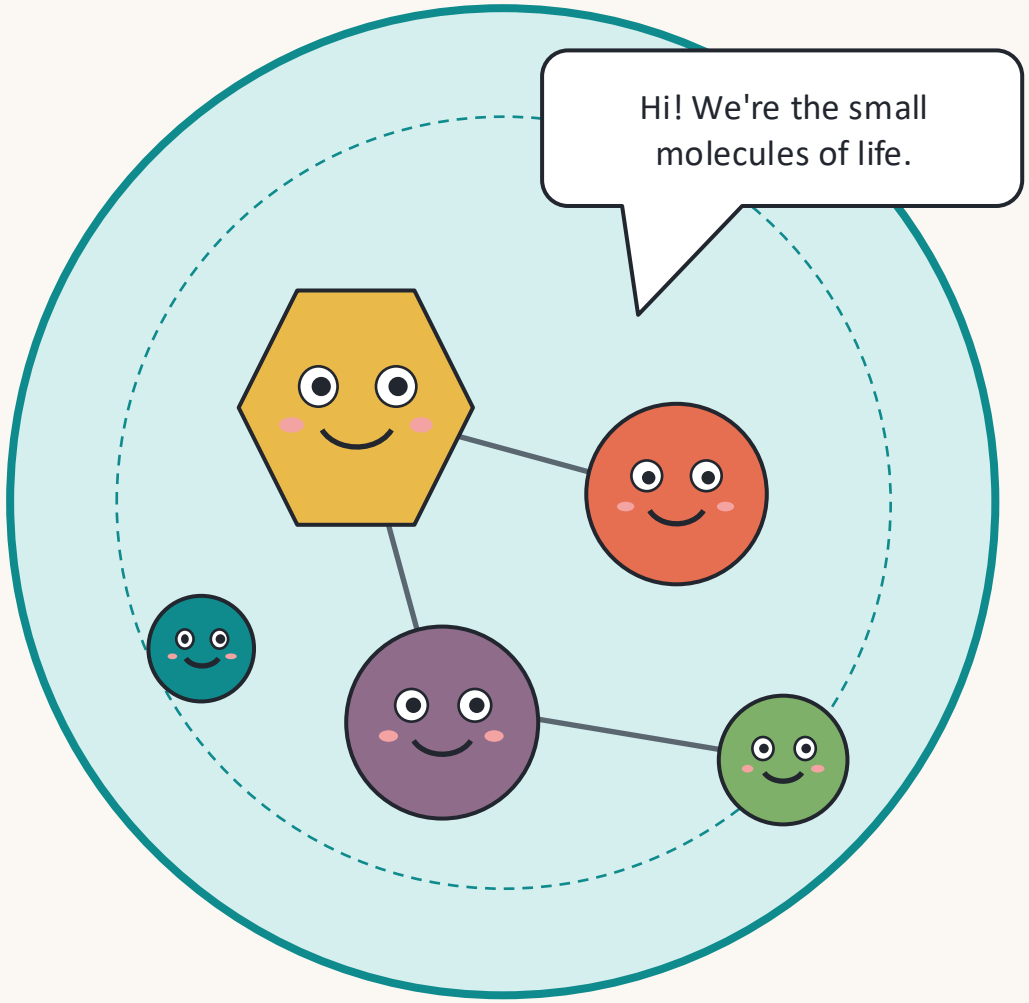
Metabolomics

Reading the chemistry of life

An introduction for students, from the basics

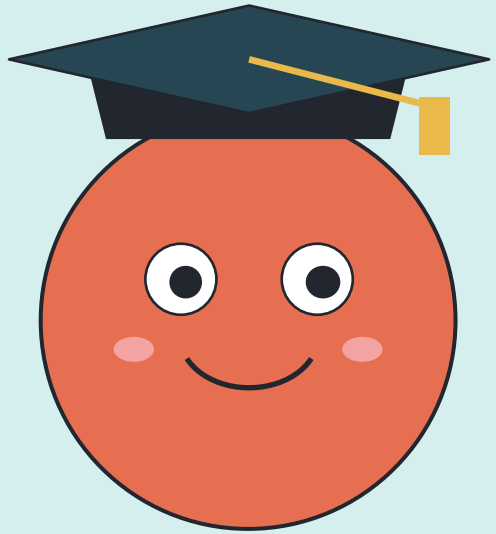
Dr. Muhammad Aammar Tufail

www.codanics.com



Sugar, amino acid, lipid, nucleotide and vitamin, drawn as cartoons

Learning goals

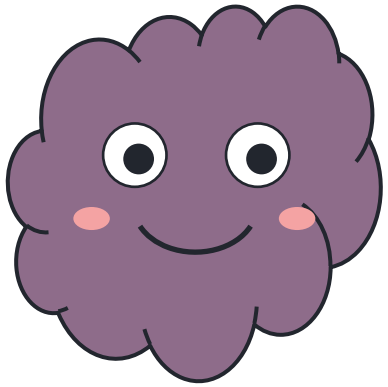


By the end of class you can…

1. Define **metabolites** and the metabolome
2. Place metabolomics among the **other omics**
3. Follow a study from **sample to biology**
4. Explain how **MS and NMR** detect metabolites
5. Read **spectra, PCA, volcano plots** and pathway maps
6. Name real **applications** and open challenges

Metabolites are the small molecules a cell makes, uses or takes in

How small is “small”?

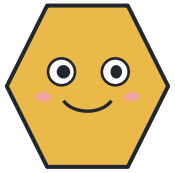


Enzyme (protein)
~50,000 Da

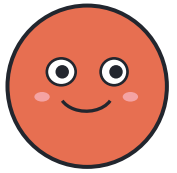


Glucose
180 Da

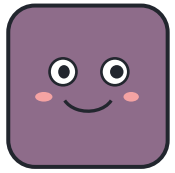
- **Metabolites** are small molecules, usually < 1,500 Da
- They are the inputs, intermediates and outputs of the cell’s chemical reactions



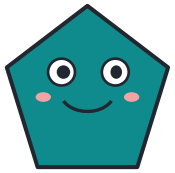
Sugars
glucose, fructose



Amino acids
glutamine, leucine



Lipids
cholesterol, fatty acids



Nucleotides
ATP, NAD^+



Organic acids
citrate, lactate



Foreign (xenobiotics)
caffeine, drugs

Source: Wishart et al., HMDB 5.0, Nucleic Acids Research 2022; Liu & Locasale, Trends in Biochemical Sciences 2017. Protein size is a typical value.

The metabolome is the omics layer closest to the phenotype



Genome

DNA

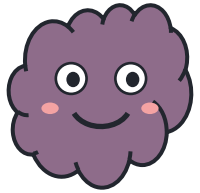
"What can happen"



Transcriptome

RNA

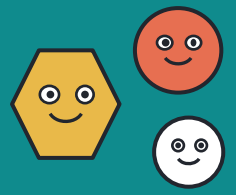
"What the cell plans to do"



Proteome

Proteins

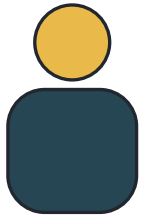
"What does the work"



Metabolome

Metabolites

"What is happening now"



Phenotype

Traits & health

"What we observe"

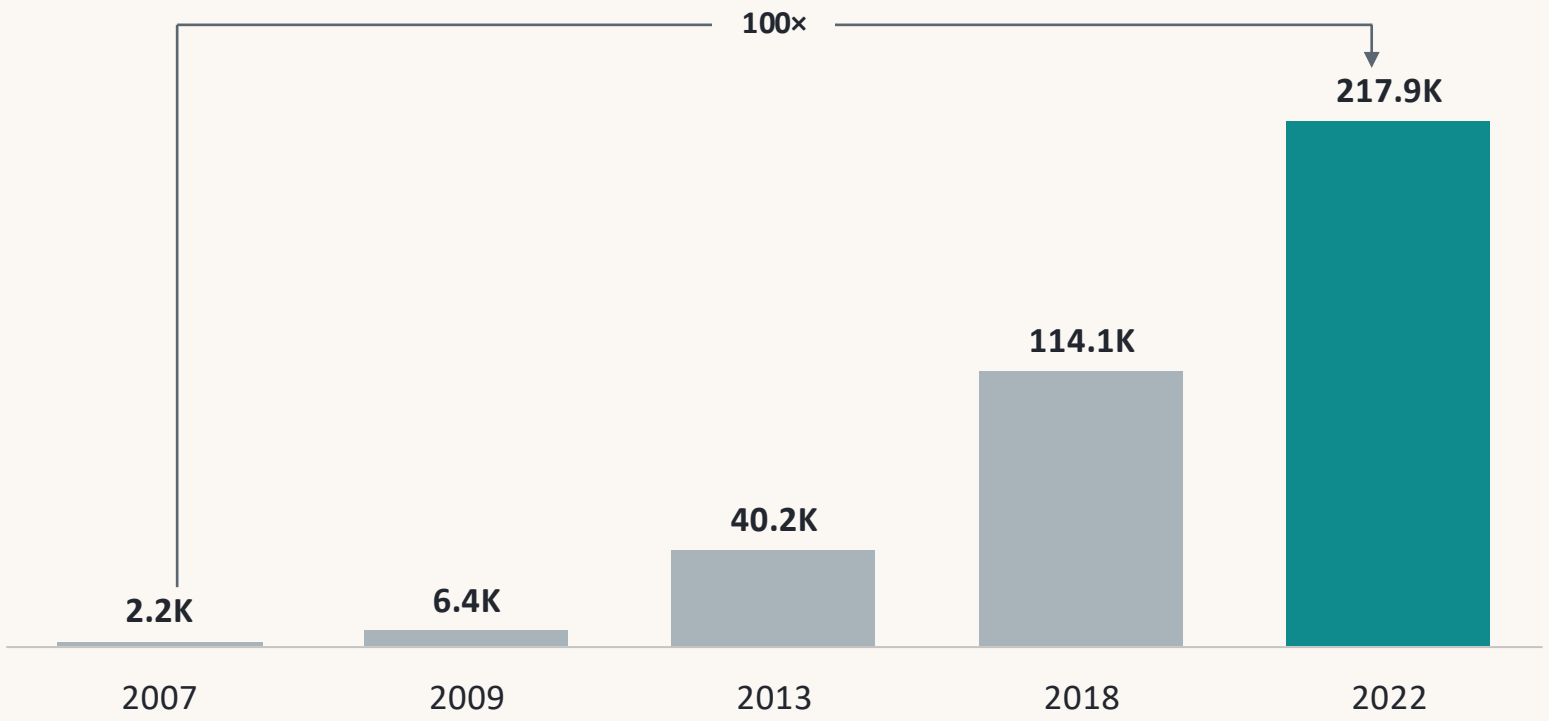
Genes say what *could* happen. Metabolites show the chemistry **going on right now**, so they track the phenotype most closely.

Diet, drugs, gut microbes and environment change metabolites directly

Source: Fiehn, Plant Molecular Biology 2002 ("the link between genotypes and phenotypes"); Patti, Yanes & Siuzdak, Nat Rev Mol Cell Biol 2012.

The human metabolome catalogue grew 100× in 15 years

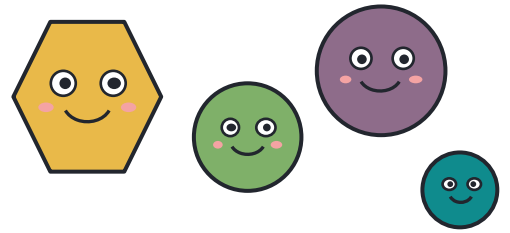
Metabolite entries in the Human Metabolome Database, by release year



<1%

of known human metabolites are identified in a typical study, which rarely names more than 900

Lots of us are still unnamed!



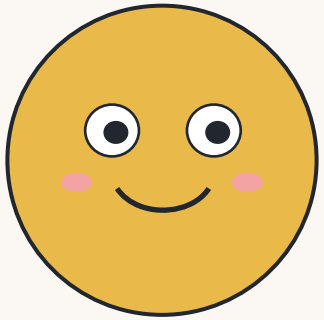
Source: Wishart et al., “HMDB 5.0: the Human Metabolome Database for 2022,” Nucleic Acids Research 50:D622–D631 (2022).

Your metabolome mixes chemistry from five sources

Our own cells
glucose, lactate, cholesterol

Food
vitamins, plant polyphenols

Gut microbes
short-chain fatty acids, TMAO



You
one metabolome

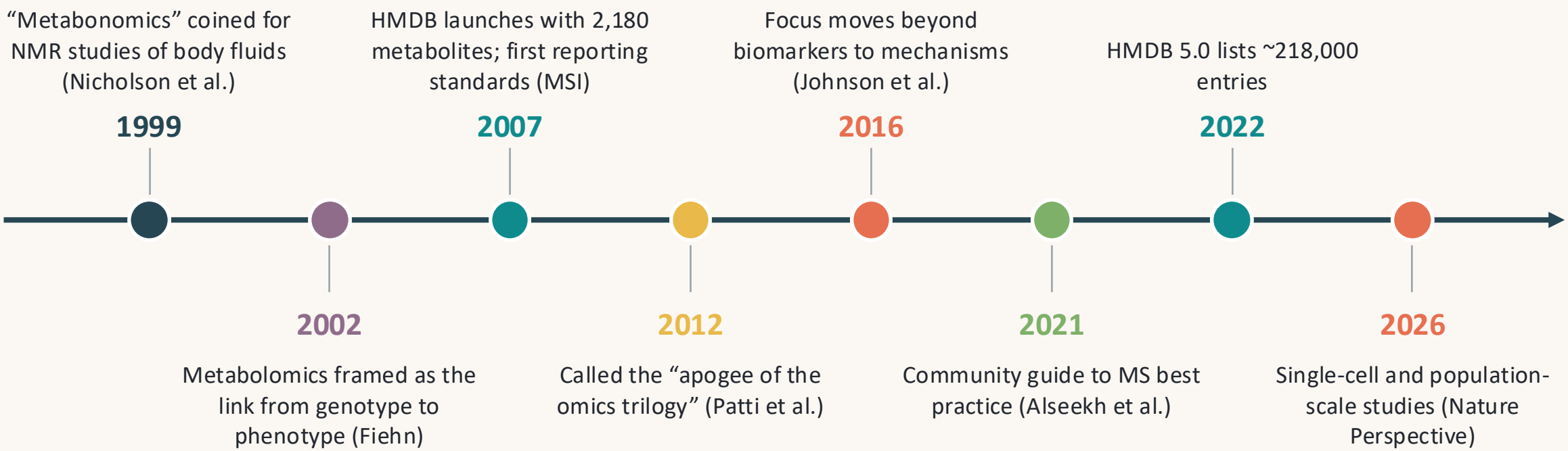
Drugs
paracetamol and its breakdown products

Environment
pollutants, pesticides, plastics

Exposome: everything outside your genes that you are exposed to over a lifetime

Source: Wishart et al., HMDB 5.0 (2022), which catalogues endogenous, food, microbial and drug-derived metabolites; Johnson, Ivanisevic & Siuzdak, Nat Rev Mol Cell Biol 2016.

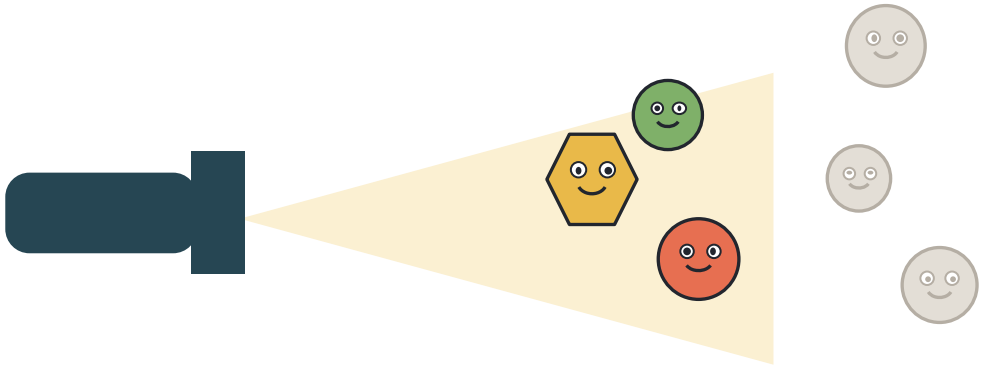
Metabolomics went mainstream in about 25 years



Sources: Nicholson, Lindon & Holmes, Xenobiotica 1999; Fiehn, Plant Mol Biol 2002; Sumner et al., Metabolomics 2007; Patti et al., NRMCB 2012; Johnson et al., NRMCB 2016; Alseekh et al., Nat Methods 2021; Wishart et al., NAR 2022; Nature 2026.

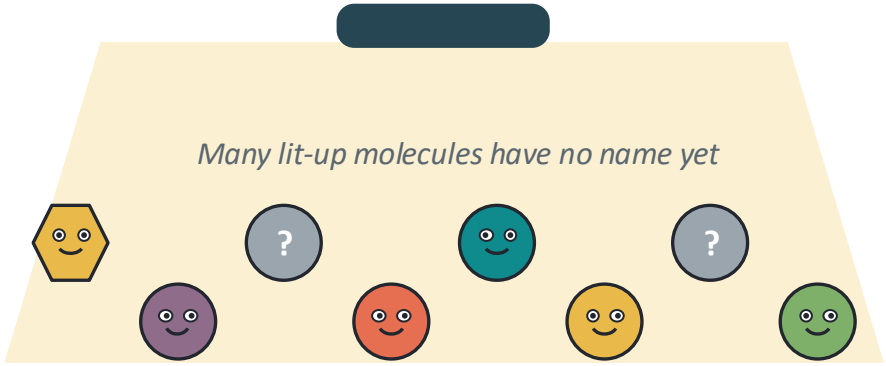
Targeted tests a known list; untargeted scans widely

Targeted



- Tens to a few hundred **known** metabolites
- **Exact amounts**, using pure standards
- Best for **testing** a hypothesis

Untargeted



- Thousands of signals, **many unknown**
- **Relative amounts** (fold changes)
- Best for **discovering** new hypotheses

Many labs do both: discover with untargeted, then confirm with targeted

Source: Patti et al., NRMCB 2012; Alseekh et al., Nat Methods 2021.

Every metabolomics study follows the same eight steps

1 PLAN

Ask a question

e.g. does drug X change liver metabolism?

2 PLAN

Design the study

groups, sample size, random run order

3 WET LAB

Collect & quench

freeze fast so enzymes stop working

4 WET LAB

Extract

solvents pull metabolites out of the sample

5 WET LAB

Measure

LC-MS, GC-MS or NMR

6 DRY LAB

Process data

raw signals become a feature table

7 DRY LAB

Analyze & identify

statistics, then name the molecules

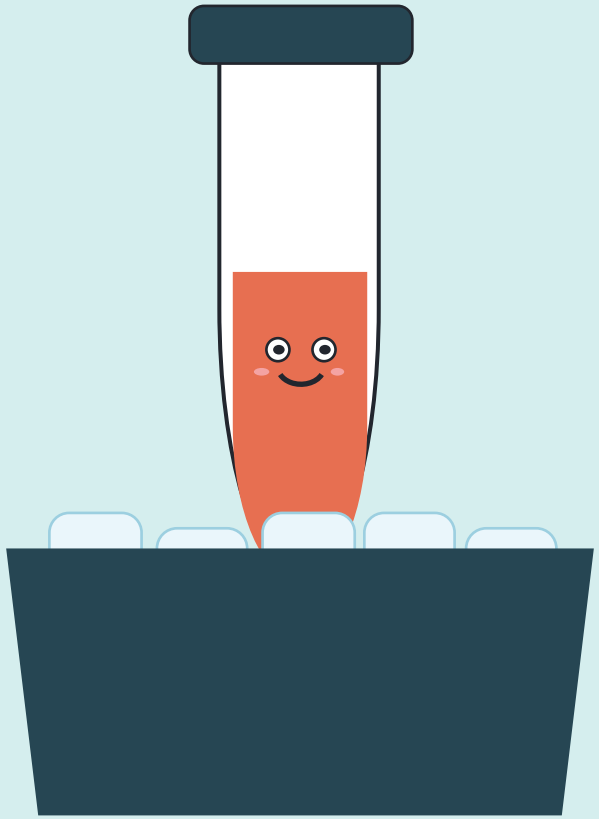
8 DRY LAB

Interpret

map to pathways; form new hypotheses

Source: workflow adapted from Johnson et al., NRMCB 2016 and Alseekh et al., Nat Methods 2021.

Data quality is decided before the instrument runs



Freeze it fast! Enzymes keep working after sampling.

1. **Biological replicates**: in every group, not only repeat injections of one sample
2. **Randomized run order**: so instrument drift cannot pose as biology
3. **Fast quenching**: liquid nitrogen or cold solvent stops metabolism
4. **Pooled QC samples**: run throughout the batch to track drift
5. **Internal standards**: check recovery and correct for ion suppression
6. **Blanks**: reveal contamination from tubes and solvents

Source: Alseekh et al., “Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices,” Nature Methods 18:747–756 (2021).

Mass spectrometry weighs molecules as charged ions

1. Separate

LC or GC column



Molecules exit the column at different times

2. Ionize

electrospray (ESI)



Each molecule gains or loses a proton and becomes an ion

3. Sort

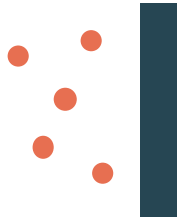
mass analyzer



Ions are sorted by mass-to-charge (m/z)

4. Detect

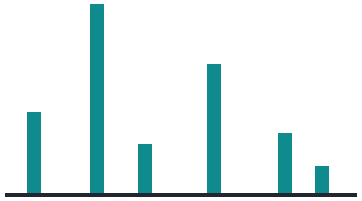
detector



Ions are counted at each m/z

5. Spectrum

computer



Each peak is an ion; its height reflects amount

Key idea: m/z = mass ÷ charge. Most metabolite ions carry a single charge, so m/z is the molecule's mass plus or minus a proton (or an adduct like Na^+).

Example: glucose (180.063 Da) appears at m/z 179.056 as $[M-H]^-$ or at m/z 203.053 as $[M+Na]^+$.

Source: Patti, Yanes & Siuzdak, NRMCB 2012; Alseekh et al., Nat Methods 2021. Masses are monoisotopic values from HMDB.

Fragment patterns (MS/MS) fingerprint a molecule

1. Pick one ion
phenylalanine, m/z 166.09

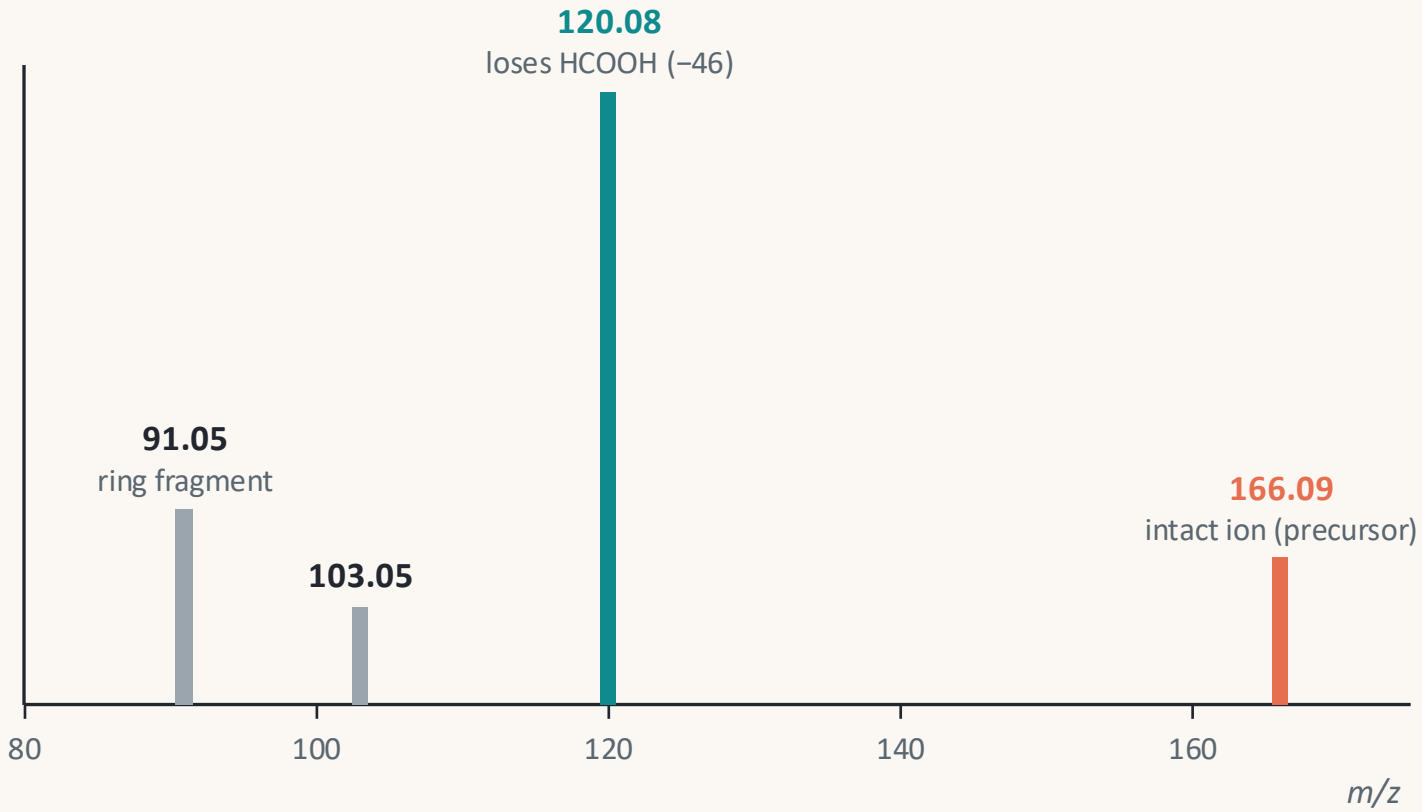
2. Smash it
collide it with gas (CID)

3. Record fragments
their masses form the MS/MS spectrum

4. Match a library
e.g. HMDB, MoNA, GNPS

Same fragments + same mass = strong evidence of identity

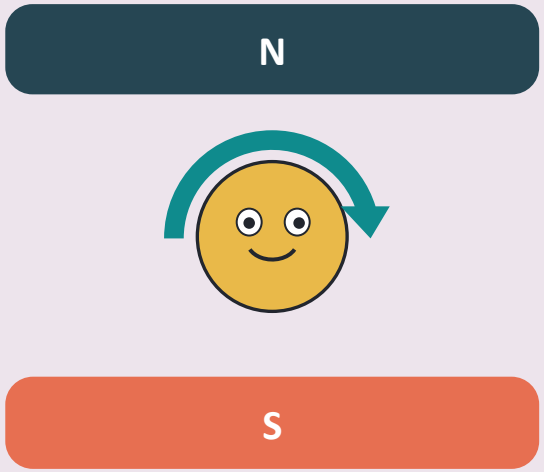
MS/MS spectrum of phenylalanine $[M+H]^+$, relative intensity (illustrative)



Fragment masses: phenylalanine reference MS/MS spectra in HMDB and MassBank of North America (MoNA). Intensities simplified for teaching.

MS detects more molecules; NMR is more reproducible

How NMR works



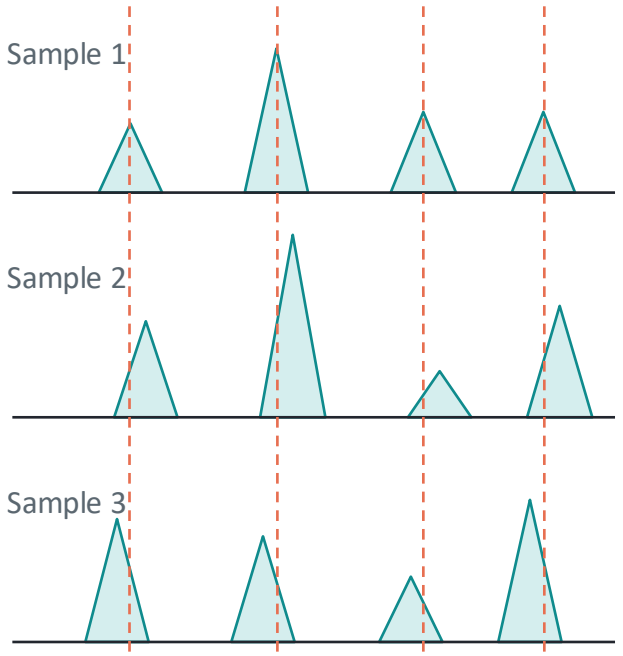
Atomic nuclei (often ^{1}H) in a strong magnet absorb radio waves. The frequency depends on each atom's chemical neighbors.

	Mass spectrometry	NMR
Sensitivity	Very high (nM or lower)	Moderate (µM)
Metabolites per run	Hundreds to thousands	Tens to ~100
Sample after analysis	Consumed	Recovered intact
Quantification	Needs standards	Direct: signal ∝ amount
Reproducibility	Varies by lab and instrument	Very high
Structure clues	Mass and fragments	Atom-by-atom connections

Typical values. Source: Emwas et al., Metabolites 2019; Patti et al., NRMCB 2012; Alseekh et al., Nat Methods 2021.

Software turns raw signals into a table of features

Raw data: one trace per sample



Same molecule, slightly different retention time in each run

Peak picking
find every signal

Alignment
match peaks across samples

Cleaning
fill gaps, drop noise and blanks

Normalization
correct for drift and dilution

Feature table (illustrative)

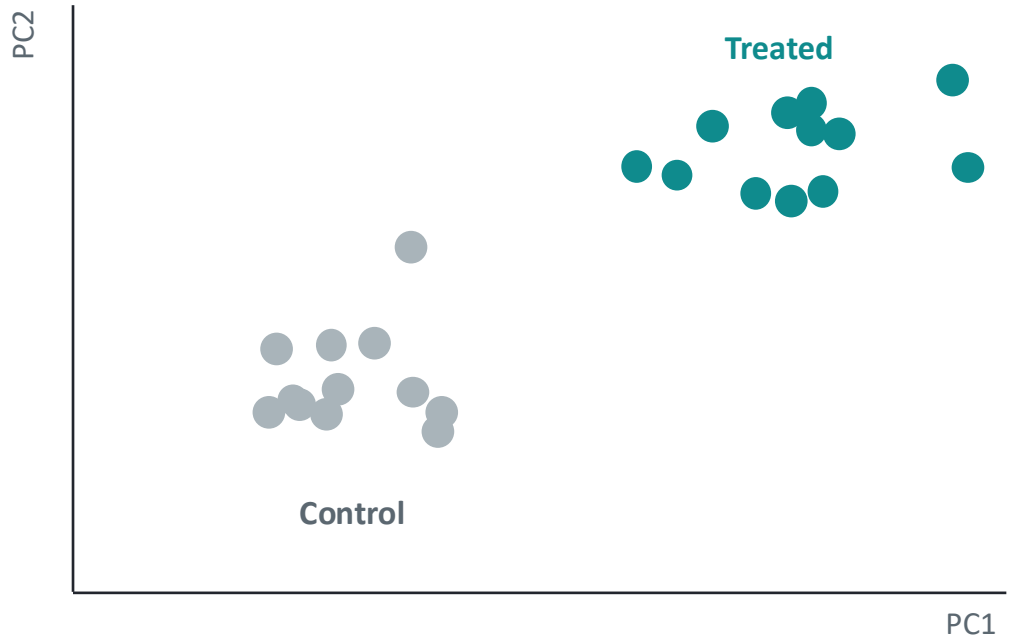
	m/z 166.09 3.2 min	**m/z 180.06** 1.1 min	**m/z 203.05** 1.1 min
Control 1	300	500	400
Control 2	350	550	450
Treated 1	900	500	200
Treated 2	850	450	150

A feature is one m/z at one retention time. It is a signal, not yet a named metabolite.

Common free software: XCMS, MZmine, MS-DIAL. Source: Alseekh et al., Nat Methods 2021; Johnson et al., NRMCB 2016.

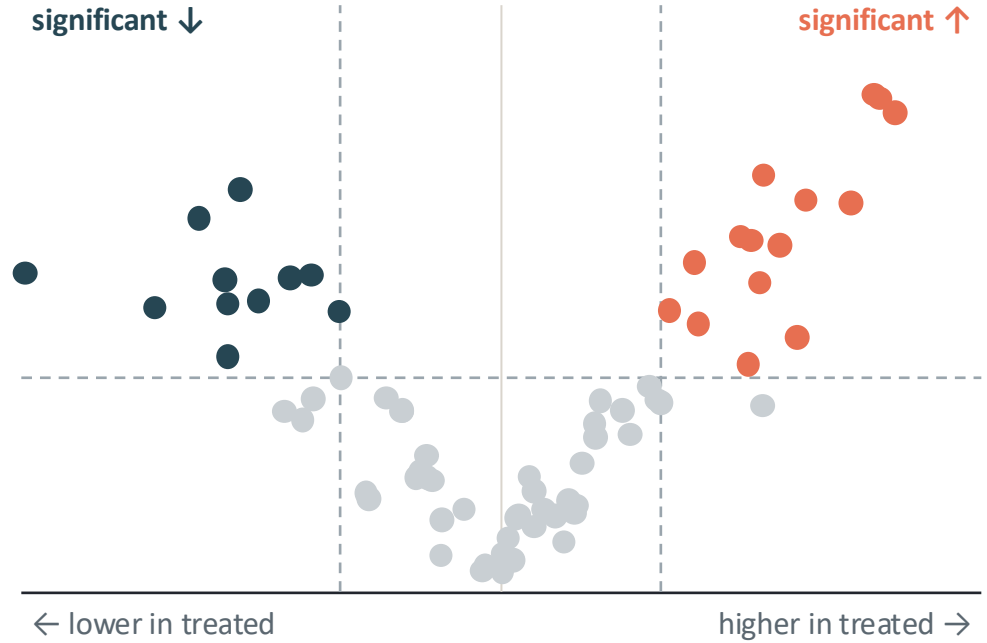
Statistics find the features that separate the groups

PCA: a map of all samples



Each dot is a sample. Samples with similar metabolite profiles sit close together.

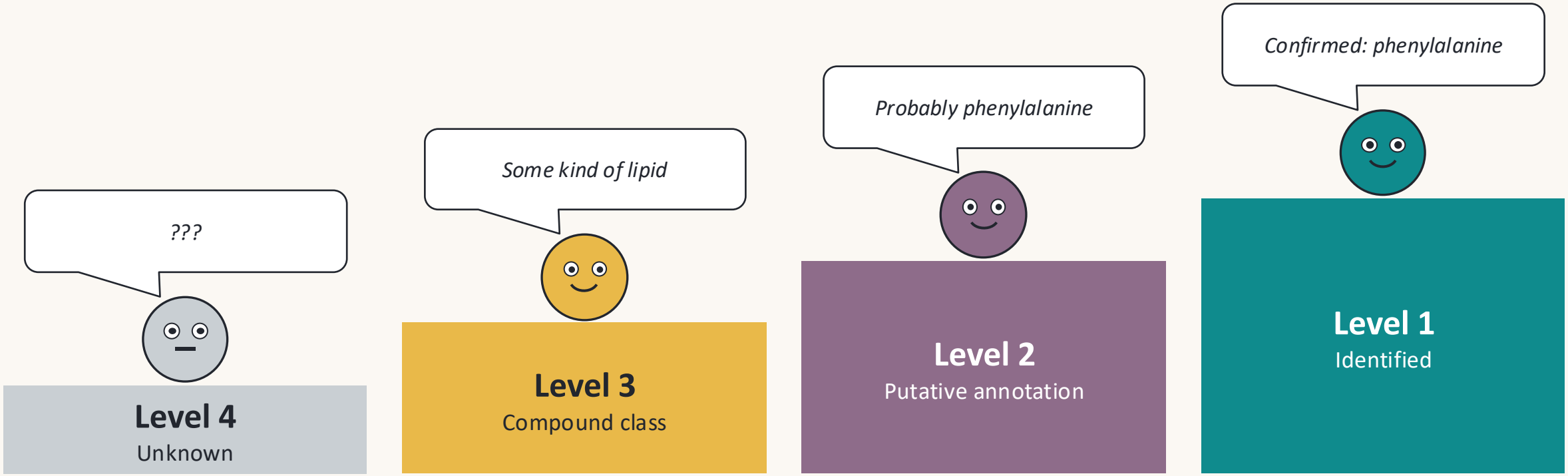
Volcano plot: which features changed?



Each dot is a feature. Far left or right = big change; high up = statistically reliable.

Illustrative data. Free tool for these analyses: MetaboAnalyst (Pang et al., Nucleic Acids Research 2024). Source: Johnson et al., NRMCB 2016.

Naming a metabolite comes with a confidence level



Seen again and again, but no identity yet

Only the chemical family is known

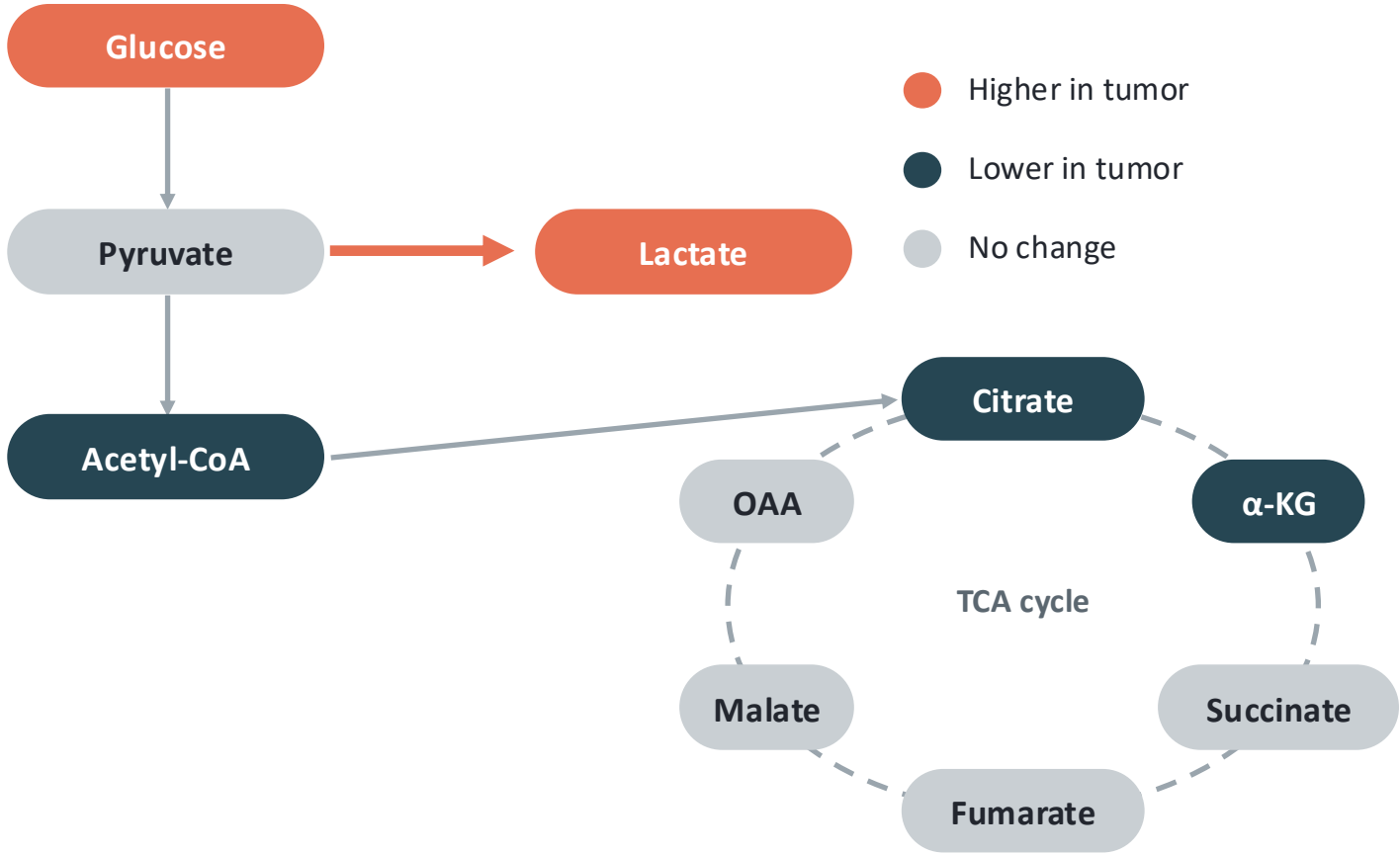
Spectrum matches a library, no standard run

Matches a pure standard on two properties (e.g. retention time and MS/MS)

Source: Sumner et al., “Proposed minimum reporting standards for chemical analysis,” Metabolomics 3:211–221 (2007), Metabolomics Standards Initiative; Alseekh et al., Nat Methods 2021.

Pathway maps turn changed molecules into biology

Example: tumor vs normal tissue (simplified, illustrative)



How pathway analysis works

1. Map changed metabolites onto known pathways (KEGG, Reactome, SMPDB)
2. Test whether a pathway holds more hits than chance (enrichment)
3. Look for where flow is blocked or rerouted

Here: glucose is pushed to lactate even with oxygen present, the Warburg effect seen in many cancers.

Source: Johnson et al., NRMCB 2016; Martínez-Reyes & Chandel, Nat Rev Cancer 2021; enrichment tools in MetaboAnalyst (Pang et al., NAR 2024). Changes shown are illustrative.

Metabolomics is already used from hospitals to farms

Newborn screening

A heel-prick blood spot is screened by tandem MS for dozens of inherited disorders, such as PKU (high phenylalanine).

Diabetes risk

Five blood amino acids flagged who developed type 2 diabetes up to 12 years later (Wang et al., 2011).

Cancer

Metabolomics revealed 2-hydroxyglutarate, an “oncometabolite” made by mutant IDH1 in brain tumors (Dang et al., 2009).

Population health

NMR profiles of ~118,000 UK Biobank adults predicted the risk of 24 diseases (Buergel et al., 2022).

Gut microbes & heart

TMAO, made from diet by gut microbes, was linked to heart disease risk (Wang et al., 2011).

Crops & food

Mapping flavor chemicals guided breeding of tastier tomatoes (Tieman et al., 2017).

Sources: Wang TJ et al., Nat Med 2011; Dang et al., Nature 2009; Buergel et al., Nat Med 2022; Wang Z et al., Nature 2011; Tieman et al., Science 2017; Wishart, Nat Rev Drug Discov 2016.

Unnamed molecules are the field’s biggest open problem

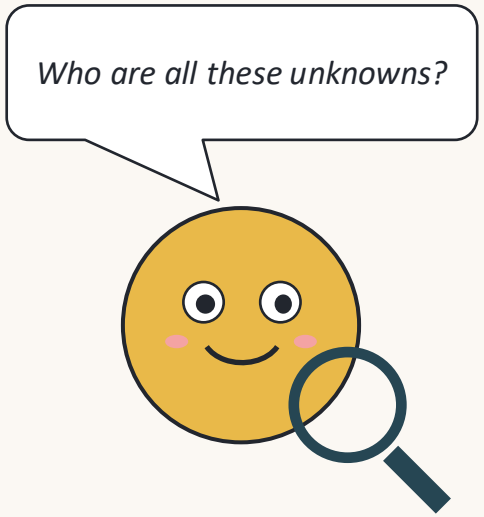
Open problems

Unknowns
most detected features still have no name

Standards
results vary between labs and instruments

Speed
metabolites change within seconds of sampling

Cause or effect?
a changed metabolite may be a result, not a driver



next tools

Frontiers

Single-cell
measure metabolites cell by cell

Spatial imaging
map molecules across a tissue slice

Population scale
profile blood from 100,000+ people

AI annotation
software predicts structures from MS/MS

Source: Wishart et al., NAR 2022; Alseekh et al., Nat Methods 2021; “Metabolomics across scales: from single cells to population studies,” Nature 2026; Dührkop et al., SIRIUS 4, Nat Methods 2019.

Key papers and resources

Start here: reviews

Liu & Locasale (2017) Metabolomics: a primer. *Trends Biochem Sci 42:274–284*

Patti, Yanes & Siuzdak (2012) Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol 13:263–269*

Johnson, Ivanisevic & Siuzdak (2016) Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol 17:451–459*

Rinschen et al. (2019) Identification of bioactive metabolites using activity metabolomics. *Nat Rev Mol Cell Biol 20:353–367*

Wishart (2016) Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov 15:473–484*

Methods, standards and databases

Alseekh et al. (2021) Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods 18:747–756*

Sumner et al. (2007) Proposed minimum reporting standards for chemical analysis. *Metabolomics 3:211–221*

Emwas et al. (2019) NMR spectroscopy for metabolomics research. *Metabolites 9:123*

Wishart et al. (2022) HMDB 5.0: the Human Metabolome Database for 2022 (hmdb.ca). *Nucleic Acids Res 50:D622–D631*

Pang et al. (2024) MetaboAnalyst 6.0, free online analysis (metaboanalyst.ca). *Nucleic Acids Res 52:W398–W406*