

# Metabolites and metabolomics

---

Data generation and pre-processing



University of  
**BRISTOL**



# Metabolomics

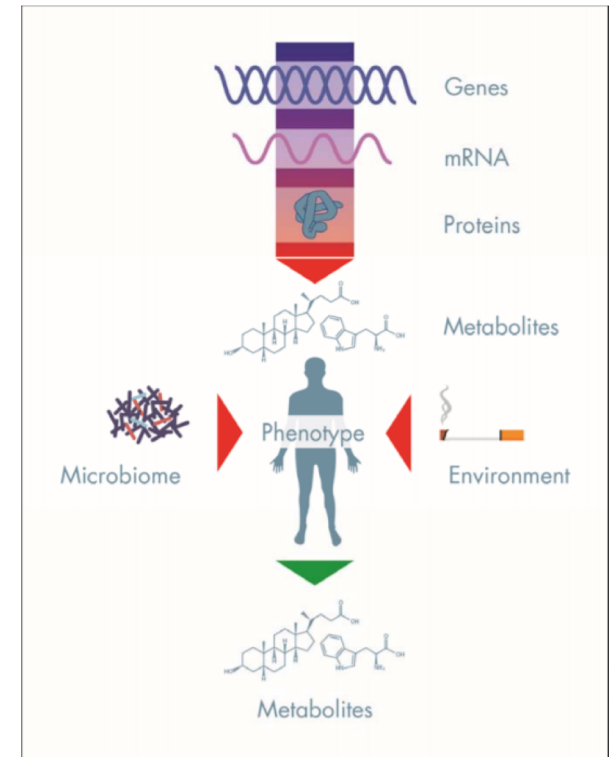
Metabolites are produced in the body as the result of chemical processes

They give a fine-grained snapshot of a person's metabolic state – offering molecular insight on health and disease

*Peptides, oligonucleotides, sugars, organic acids, ketones, aldehydes, amines, amino acids, lipids, steroids, and alkaloids*

No universal instrument or method capable of measuring the entire metabolome (yet...)

Two main types: nuclear magnetic resonance (NMR) and mass spectrometry (MS)



<https://www.metabolon.com/application/files/2415/0388/1050/The-5-Key-Elements-of-a-Successful-Metabolomics-Study.pdf>

# NMR and MS

NMR and MS have distinct strengths and weaknesses, and both uniquely benefit the study of metabolomics.

NMR	MS
Quantitative	Higher resolution
Less sensitive (but improving!)	More sensitive
Cheaper	More expensive
Can only detect most abundant metabolites ~200	More metabolites can be detected ~1200
Non-destructive (sample can be used again!)	Destructive
Highly reproducible	Becoming more widely used

Rarely done in epidemiology, but by combining the approaches you can enhance coverage of the metabolome and the accuracy of metabolite identification



University of  
BRISTOL



# Examples



- Nightingale NMR output
- Concentrations for each NMR-derived metabolite, measured in mmol/L

	metabolnmrsampleid	xxlvidlp	xxlvidll	xxlvidlpl	xxlvidlc	xxlvidlce	xxlvidlfc	xxlvidltg	xlvidlp	xlvidll	xlvidlpl
1	IEU170613.BOSEC.170612088.1174729	3.396e-11	0.0074410	8.812e-04	0.0018730	1.285e-03	5.880e-04	0.004688	1.208e-10	0.012020	0.0025390
2	IEU170613.BOSEC.170612088.1174738	4.950e-11	0.0106900	1.075e-03	0.0021140	1.103e-03	1.011e-03	0.007497	4.265e-10	0.040470	0.0055050
3	IEU170613.BOSEC.170612088.1174682	0.000e+00	0.0000000	0.000e+00	0.0000000	0.000e+00	0.000e+00	0.000000	0.000e+00	0.000000	0.0000000
4	IEU170613.BOSEC.170612088.1174673	1.106e-10	0.0239200	2.899e-03	0.0046660	2.501e-03	2.164e-03	0.016350	7.598e-10	0.073420	0.0119300
5	IEU170613.BOSEC.170612088.1174664	4.414e-11	0.0097900	1.162e-03	0.0025650	1.550e-03	1.015e-03	0.006063	3.166e-10	0.030850	0.0054140

- Metabolon MS output
- Raw area counts

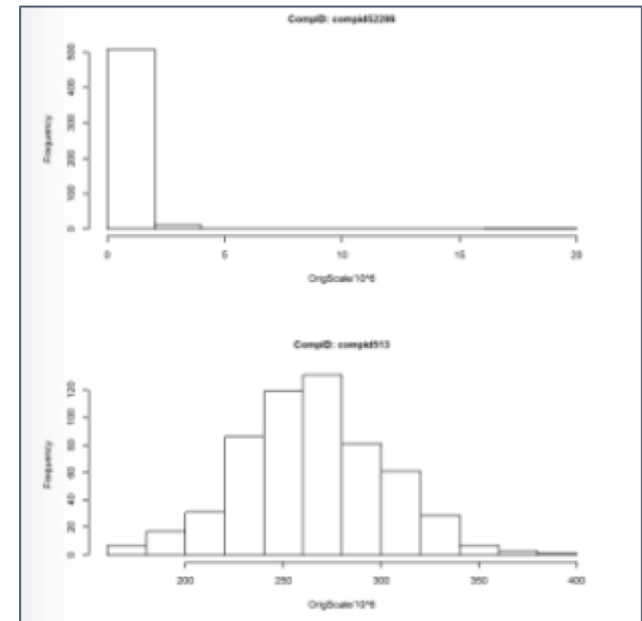
metmsr52740	metmsr52746	metmsr52743	metmsr57814	metmsr48762	metmsr52603	metmsr53174	metmsr19130	metmsr57341	metmsr34404	metmsr32391	metmsr20675	metmsr34400	metmsr44621
NA	NA	NA	1352018.5	3566296.25	192019552	2395288.8	110228472	901221.2	592798.81	316324.50	59937624	1637726.12	300206.7
NA	NA	NA	1474427.6	111167.21	581630336	1548944.4	73087856	407666.3	70758.06	147568.02	87926176	436513.69	582169.8
NA	NA	NA	982092.0	156365.19	790785024	1349588.9	92968632	455656.5	363982.69	310980.62	70000072	1211505.88	957168.4
NA	NA	NA	1146396.0	22284.42	102844880	250097.2	100711448	290085.0	NA	NA	106341648	300846.81	419749.0
NA	NA	NA	993444.6	778022.12	777507968	2278870.5	79280856	648250.4	19481.87	NA	95711576	188443.38	485463.0

- Metabolon MS output
- Scaled and imputed using multiples of the median

metmss52740	metmss52746	metmss52743	metmss57814	metmss48762	metmss52603	metmss53174	metmss19130	metmss57341	metmss34404	metmss32391	metmss20675	metmss34400	metmss44621
0.14577	0.28692	0.17802	0.36738	3.69408	1.86708	1.46402	0.95029	1.93314	1.96783	1.08430	0.87694	1.44072	0.52939
0.14577	0.28692	0.17802	0.40064	0.11515	5.65541	0.94673	0.63010	0.87445	0.23489	0.50584	1.28644	0.38400	1.02660
0.14577	0.28692	0.17802	0.26686	0.16197	7.68910	0.82488	0.80149	0.97739	1.20826	1.06598	1.02416	1.06577	1.68780
0.14577	0.28692	0.17802	0.31151	0.02308	1.00000	0.15286	0.86824	0.62224	0.02973	0.08228	1.55587	0.26466	0.74015
0.14577	0.28692	0.17802	0.26995	0.80590	7.56001	1.39286	0.68349	1.39051	0.06467	0.08228	1.40035	0.16578	0.85600

# What you do with your data

- **Checks for normality**
- Plotting your data e.g. histograms, QQ plots
- Statistical tests e.g. Shapiro-Wilk
- Outlier detection e.g. principal components analysis (PCA)
- **Log transform** – transforms the data
  - but will not 'fix' all distributions (often not helpful because it can't deal with 0 values)
- **Scale** – standardizes the data to z scores
- No agreed 'best practice' yet...



Example of MS-profiled metabolite distribution