IFR

Institute of
Food Research

600 MHz NMR spectrometer with cryoprobe

An introduction to metabolomics

Dr Gwénaëlle Le Gall

Norwich UK

Quadram Institute

Science Health
Food Innovation

Outline

Historic - Metabolomics

One of the first publications on metabolic profiling using high-resolution ^{1}H NMR (1984)

^{1}H NMR spectrum of a urine extract

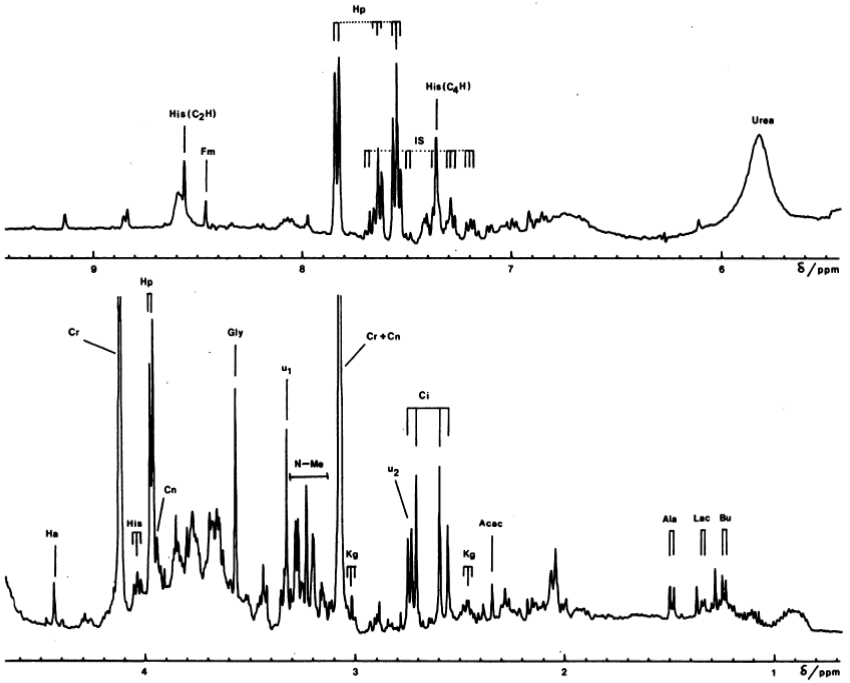


Fig. 1. 400 MHz ^{1}H NMR spectrum of normal human urine containing 100 mL $^{2}H_2O$ per liter

Upper: aromatic region; *Lower,* aliphatic region. These spectra are the result of 232 accumulations at 24 °C, using a 60 ° pulse, 9.7-s interval between pulses, and 16 384 data points. The water resonance was suppressed by continuous irradiation. An exponential function corresponding to a 1-Hz line broadening was applied. Abbreviations for peak assignments are given in Table 1; N-Me peaks are probably methyl resonances from molecules such as ergothioneine, betaine, carnitine, and choline; u = unassigned resonances

CLINICAL CHEMISTRY, Vol. 30, No. 3, 1984 427

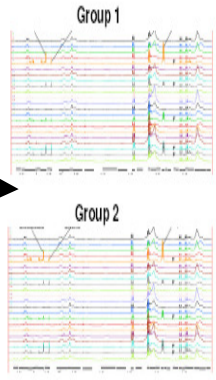
Jeremy Nicholson from Imperial college, London, UK pioneered the use of NMR to analyse metabolites in biological systems

Definition - Metabolomics

Metabolomics = detection of many metabolite features then looking for differences in groups of samples

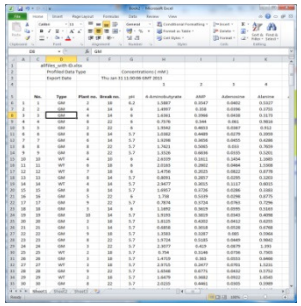
1.Samples

2. Record chemical data

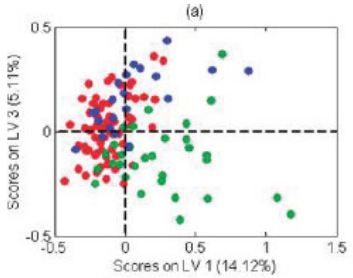


NMR = 50-70 metabolites (central metabolism)

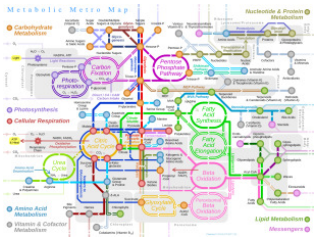
3. Process dataset



4. Analyse/Model data/Identify



5. Interpret the results

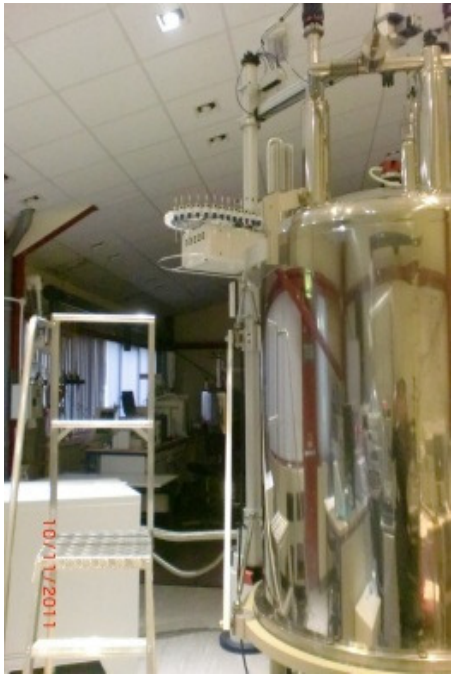


Metabolite= low molecular weight compound (< 1000 Daltons)

<u>PubMed hits:</u>
metabolomics = 14500
metabonomics = 15000
(Dec 2016)

Technical platforms

Nuclear Magnetic Resonance (NMR)



Liquid Chromatography/ Mass Spectrometry (LC/MS)

UPLC/MS-QTOF

LC/MS-TOF

LTQ Orbitrap

Untargeted Metabolomics

TOF: Time-of-Flight **Linear Trap Quadrupole**

QQQ

Targeted Metabolomics

QQQ: triple quadrupole

Gas Chromatography/ Mass Spectrometry (GC/MS) (not shown)
Capillary electrophoresis / Mass Spectrometry (CE/MS) (not shown)

Why metabolomics?

Health & wellbeing

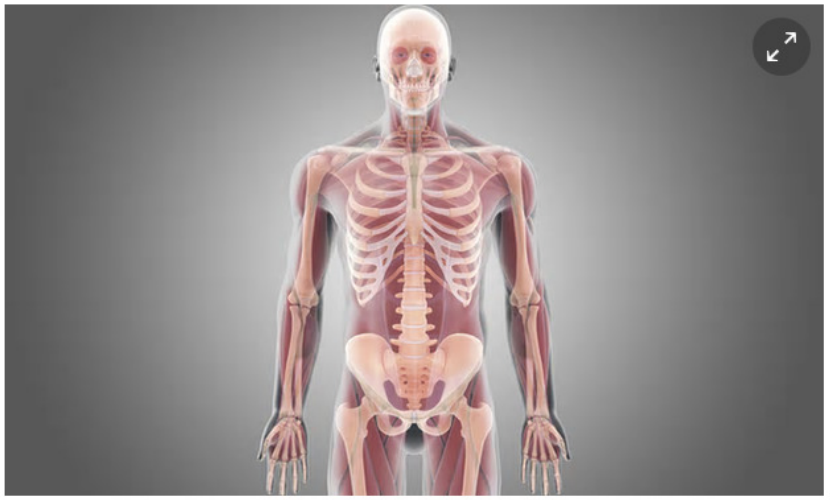
Mapping our lives: the importance of lifelong health studies

Longitudinal health studies follow subjects from 'womb to tomb' to create a fingerprint of a healthy human. And there is one thing that has a huge impact on our wellbeing

jobs dating more UK edition
theguardian
website of the year

Frankie Mullin

Sunday 24 August 2014
17.30 BST



Written on the body: genetic and molecular data is collected from thousands of individuals. Photograph: Alamy

To profile biofllluids and correlate to phenotypes

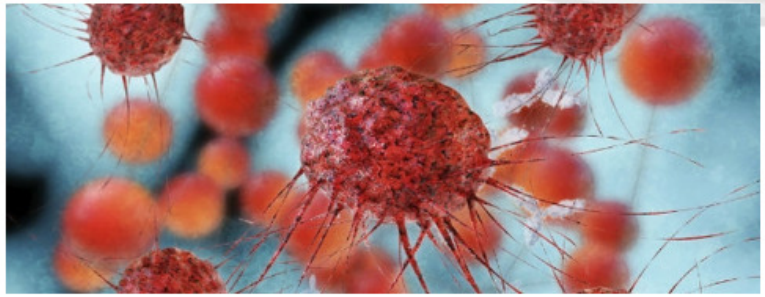
"Metabolomics [the study of cellular-level chemical processes] now allows us to get 300 measures from a sample, whereas we once got two or three. We're hoping this will allow us to better understand the biological pathways to illness. However, we still have to ask the right sort of questions."

Why metabolomics?

NEWS

MRC Medical Research Council

Leading science for better health



Metabolism 'rewiring' can lead to aggressive lung cancer

24 Feb 2016

Scientists have discovered that lung cancers with extra copies of a cancer causing gene-defect 'rewire' their energy supply, helping them to survive and making them more likely to spread.

Researchers at the MRC Cancer Unit at the University of Cambridge studied lung cancers with mutations in their Kras genes, which are found in around 30 per cent of adenocarcinomas - the most common type of primary lung cancer. They found that the number of copies of Kras mutations had a profound impact on the disease, as those with extra copies undergo a change in their metabolism.

Lung cancer cells with extra copies of Kras mutations increase their uptake of glucose - the major energy source in the body - and show alterations in the way this sugar is processed. Changes in glucose metabolism are a well-known cancer trait but this study revealed that cells with extra copies of mutant Kras utilise glucose differently from those either with a single mutation or normal lungs.

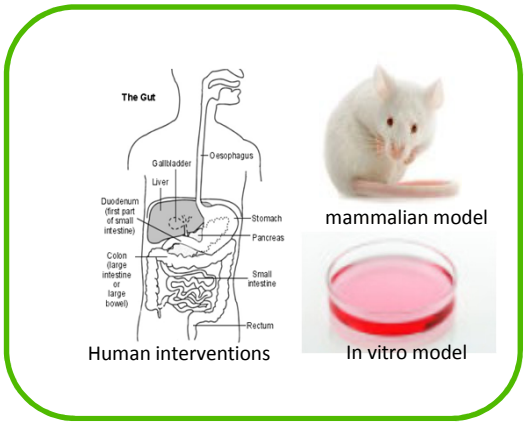
This metabolic "rewiring" enables these cells to cope better with certain cellular stresses but also means that they have unique metabolic needs that can be exploited therapeutically. Since an increase in the number of copies of Kras mutations were associated with more aggressive tumour features, such as the ability to spread, the therapeutic implications of the study are particularly appealing.

To elucidate biological mechanisms

Kerr et al, Mutant *Kras* copy number defines metabolic reprogramming and therapeutic susceptibilities, Nature (2016), 531, 110–113

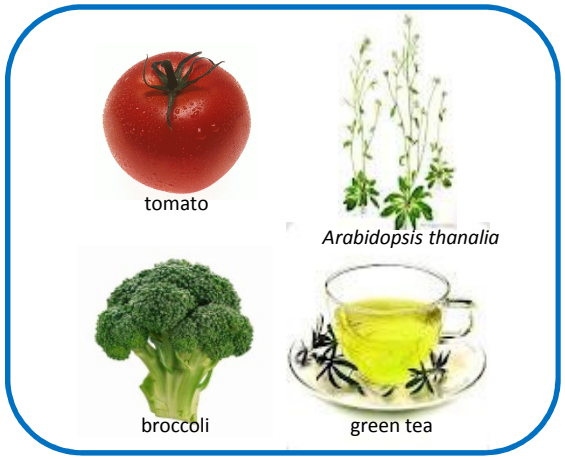
Scientific areas

Mammalian/Animal Systems



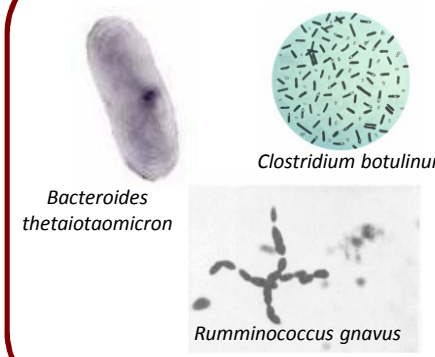
- disease/ personalized health care
- molecular epidemiology
- diet
- new drug targets

Plants / Food



- **substantial equivalence** (comparison of genetically modified/new variety versus conventional foods)

Microbes



- gene function (study of Knock-Out)
- taxonomy, species classification
- environment stimuli (stress, nutrients)
- characterising metabolism
- interaction with host

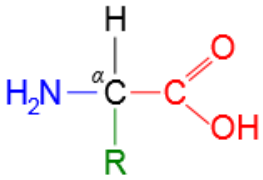
Aqueous Metabolites

NMR: detects any molecule containing ^{1}H nuclei but not sensitive

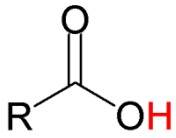
MS: low detection limit but some molecules not detected

Primary metabolites:
(all living systems)

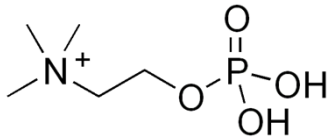
Amino acids



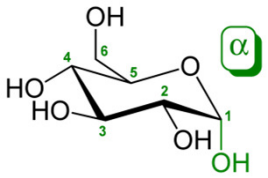
Organic acids



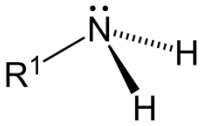
Osmolytes



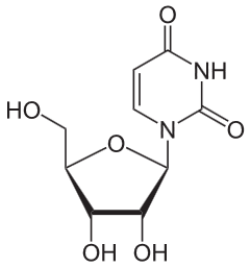
Sugars



Amines

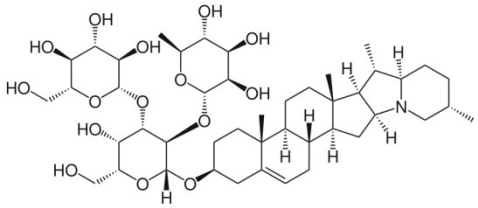


Nucleosides

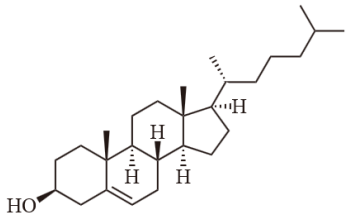


Secondary metabolites:
(plants, fungi)

Alkaloids



Steroids, Terpenoids



Phenolic compounds, Flavonoids

Lipids

NMR detects the total content of abundant lipids → measures total TG, cholesterol, PC and SM

MS detects 100s of individual lipids → screens neutral and phospholipids

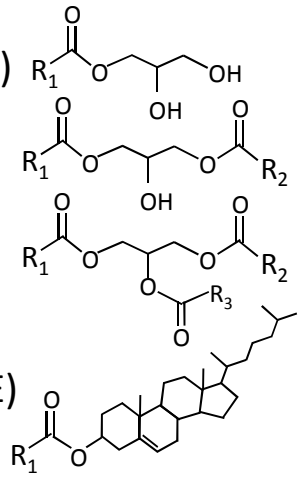
Neutral lipids

Monoglycerides (MG)

Diglycerides (DG)

Triacylglycerols (TG)

Cholesteryl esters (CE)



Main radicals R:

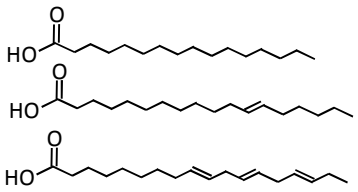
palmitic (C16:0) →

stearic (C18:0)

oleic(C18:1) →

linoleic (C18:2)

linolenic (C18:3) →



Phospholipids

Glycerophospholipids (PL) →

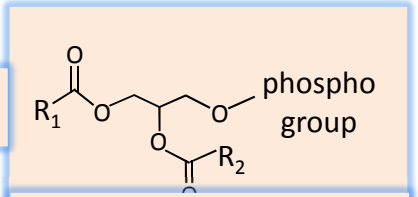
phosphatidylcholines (PC)

phosphatidylethanolamines (PE)

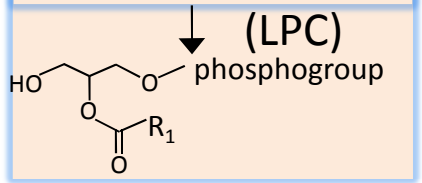
phosphatidylserines (PS)

phosphatidylinositols (PI)

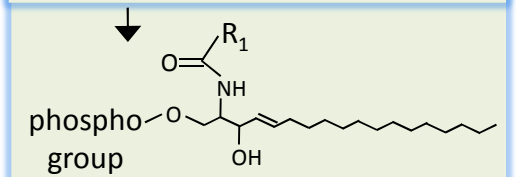
Phosphatidic acids (PA)



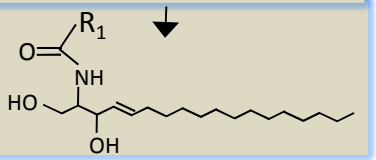
Lysophospholipids (LPC)



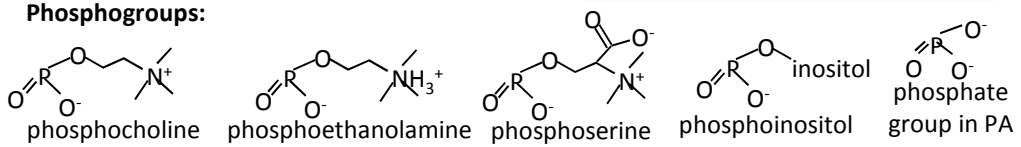
Sphingomyelins (SM)



Ceramides (CER)



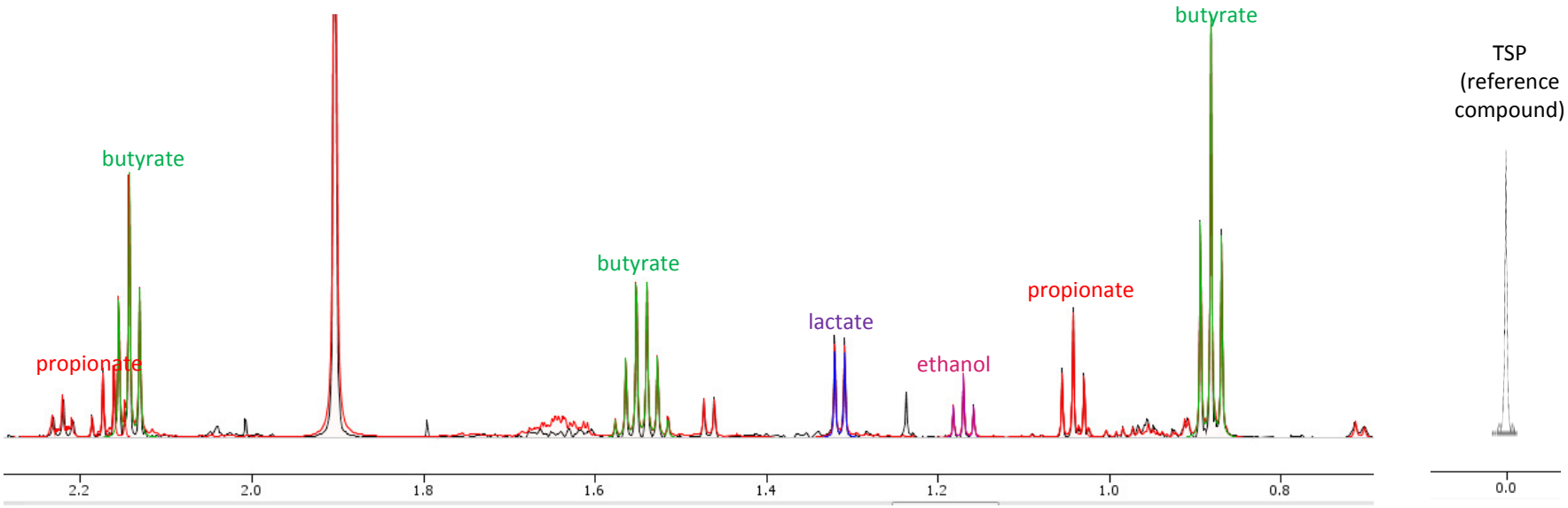
Phosphogroups:



Other lipids such as free fatty acids, arachidonic acid, prostaglandins and inositols are detected by targeted GC and LC/MS lipidomics

NMR: quantitative

^{1}H spectrum of a faecal extract in Chenomx software



Compound Name	Concentration (mM)	Maximum (mM)
Butyrate	**2.3008**	**2.2342**
Ethanol	**0.4304**	**0.4248**
Lactate	**0.4091**	**0.3921**
Propionate	**0.7498**	**0.7538**

No need for standards with ^{1}H NMR. The compound used as a reference to set the X axis, named trimethylsilyl propionate (TSP) is the unique standard used to quantify the compounds detected in the ^{1}H NMR spectrum

MS: quantitative

Meat Authentication via Multiple Reaction Monitoring Mass Spectrometry of Myoglobin Peptides

Andrew D. Watson, Yvonne Gunning, Neil M. Rigby, Mark Philo, and E. Kate Kemsley*

Analytical Sciences Unit, Institute of Food Research, Norwich Research Park, Norwich NR4 7UA, United Kingdom

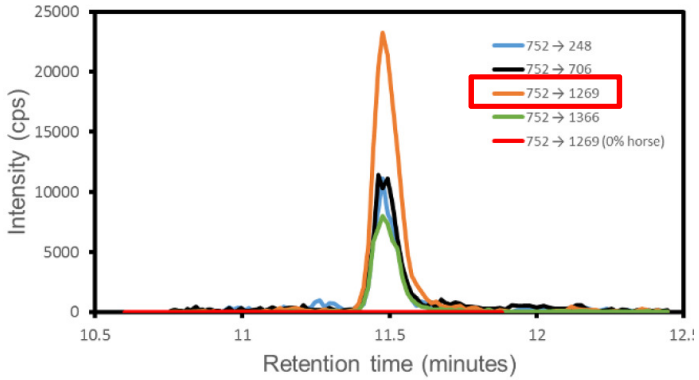


Figure 1. MRM transition intensities versus retention time for 1% w/w horse in beef. . The marker peptide is HPGDFGADAQGAMTK.
J Vis Exp. 2016, 115, e54420, 1-7

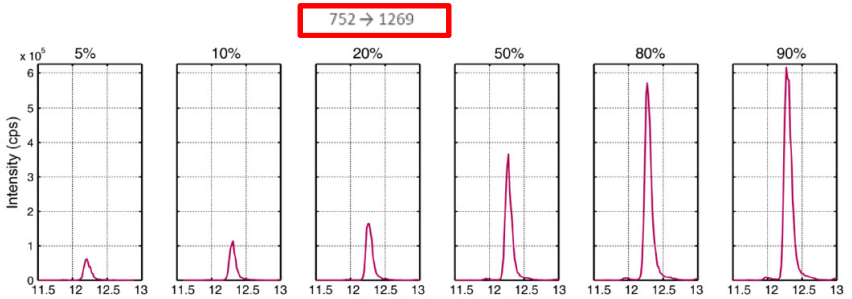


Figure 4. MRM transition intensities versus retention time for different mixtures of horse in beef, ranging from 5% to 90% horse (w/w). *Anal. Chem.* 2015, 87, 10315−10322

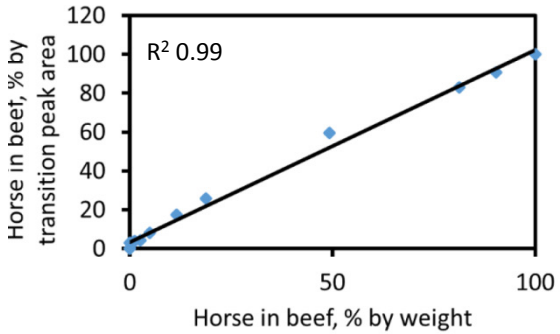
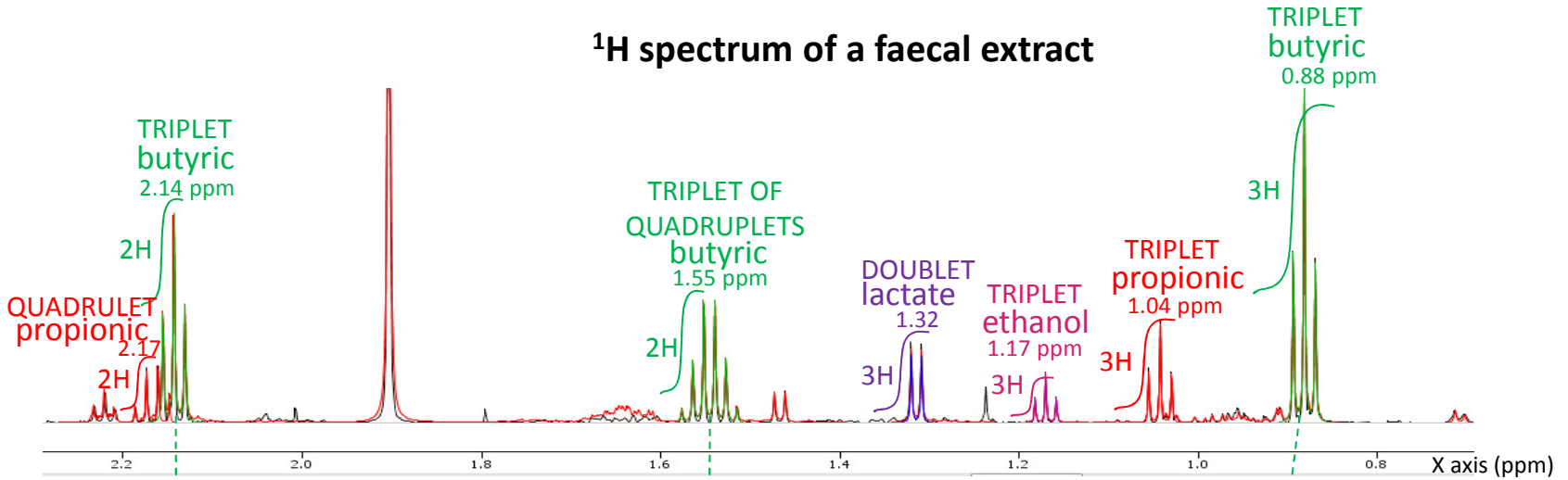
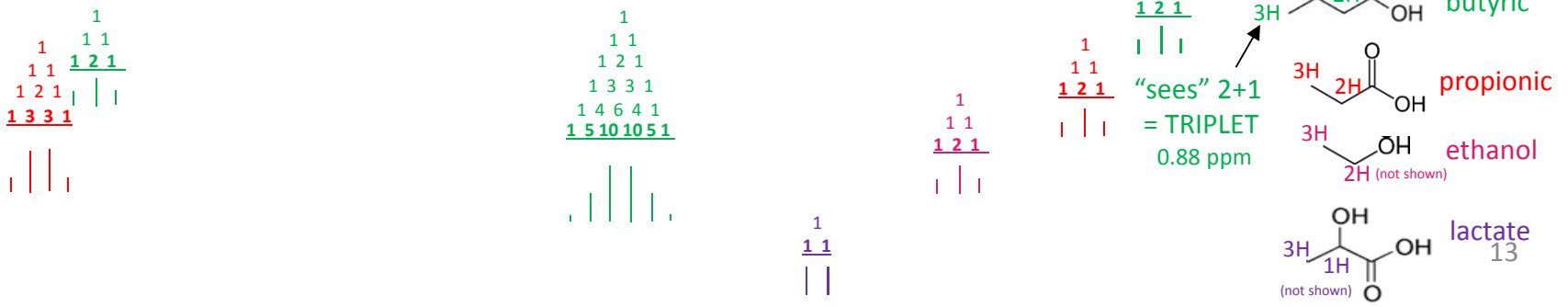


Figure 2. Plot of horse in beef, as percent weight for weight, versus horse in beef as percent transition peak area. J Vis Exp. 2016, 115, e54420, 1-7

NMR: structural elucidation



- Each compound can have one or several chemical groups or chemical shifts (location on the X axis)
- For example butyric acid has 3 chemical shifts
- Interactions with neighbouring protons result in the splitting of NMR peaks.
  Multiplicity = number of neighbours + 1
- The peak intensity distribution follows the pattern of numbers in Pascal's triangle



MS: structural elucidation

- Analysis of fragment ions

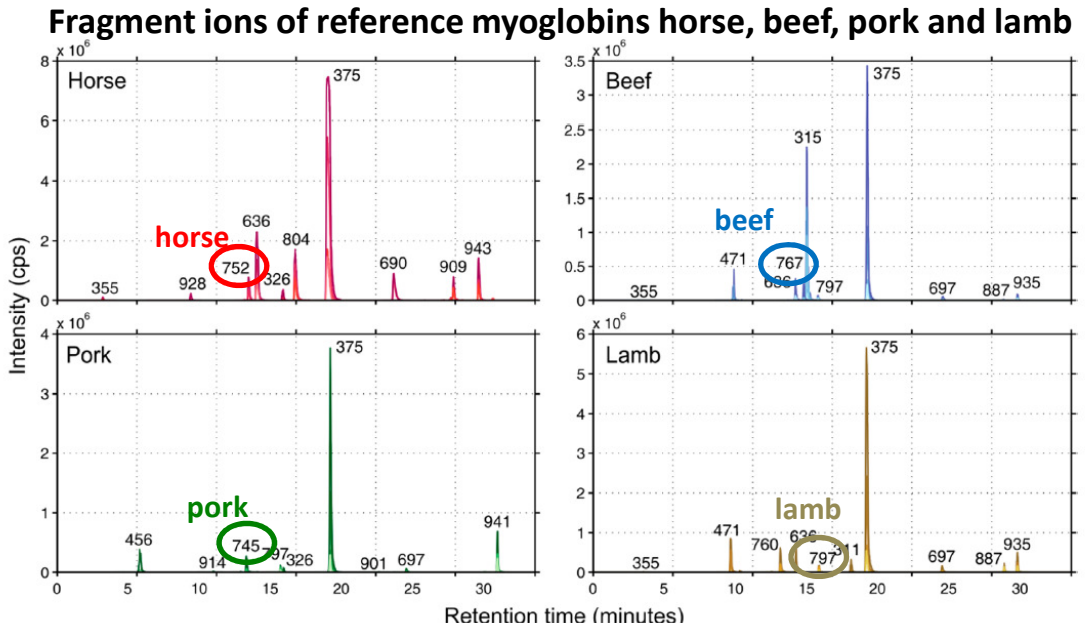


Figure 2. MRM transition intensities versus retention time for horse, beef, pork, and lamb reference myoglobins
Some peptides are shown to be common to more than one species, but others are candidates for differentiation between species.

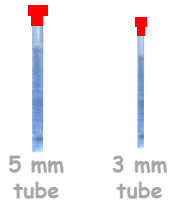
Anal. Chem. 2015, 87, 10315−10322

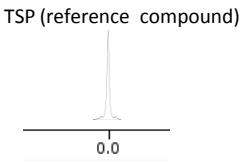
Circles: ions specific to each species

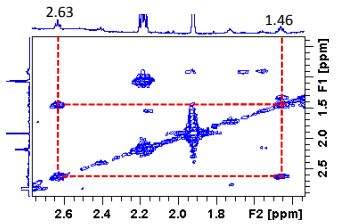
- Ion accurate mass

NMR advantages

- Simple sample preparation
- Robust automated recording (repeatable)
- High-throughput (up to 100 samples /day)
- Quantitative
- Can detect **any** metabolite of concentration above 50 μM, sometimes lower (5 μM)
- Structural elucidation of unknown compounds



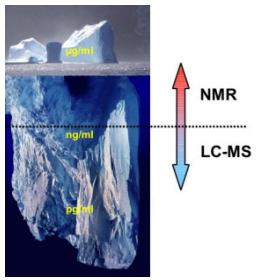




unknown triplet at 1.46 ppm linked to unknown triplet at 2.63 ppm.
Structure = Y-CH_2-CH_2-X

MS advantages

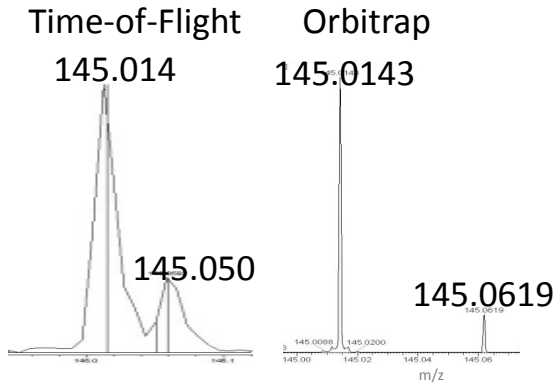
- Sensitive (down to 5 nM or lower)



- Structural elucidation of unknown compounds (fragments, accurate mass)

2-Ketoglutaric acid $C_5H_6O_5$ [MH]+ theoretical mass 145.0137

Glutamine $C_5H_{10}N2O_3$ [MH]+ theoretical mass 145.0613



- Simple sample preparation

- Automation (needs checking)

- High-throughput (up to 100 samples /day if automation is ok)

Sample extraction

Options:

Lyophilise

wine

plants, fruits, vegetables

Broccoli

Carrot

Pepper

Homogenise

meat

pestle and mortar

+

liquid nitrogen

mammalian tissue

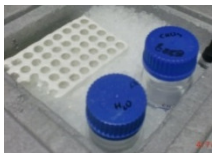
weight tissue

Keep on ice

glass beads

Homogenise

biphasic extraction with methanol, water &chloroform



Extract tissue with ice cold solvents

Dilute

urine plasma faeces

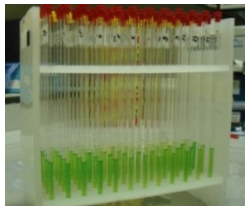
+

Phosphate buffer

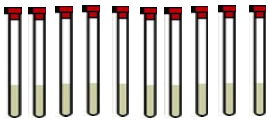
+

vortex mixer

Example:
a set of Arabidopsis extracts in NMR tubes in a rack



(dry)
Add (reconstitute in) deuterated solvent



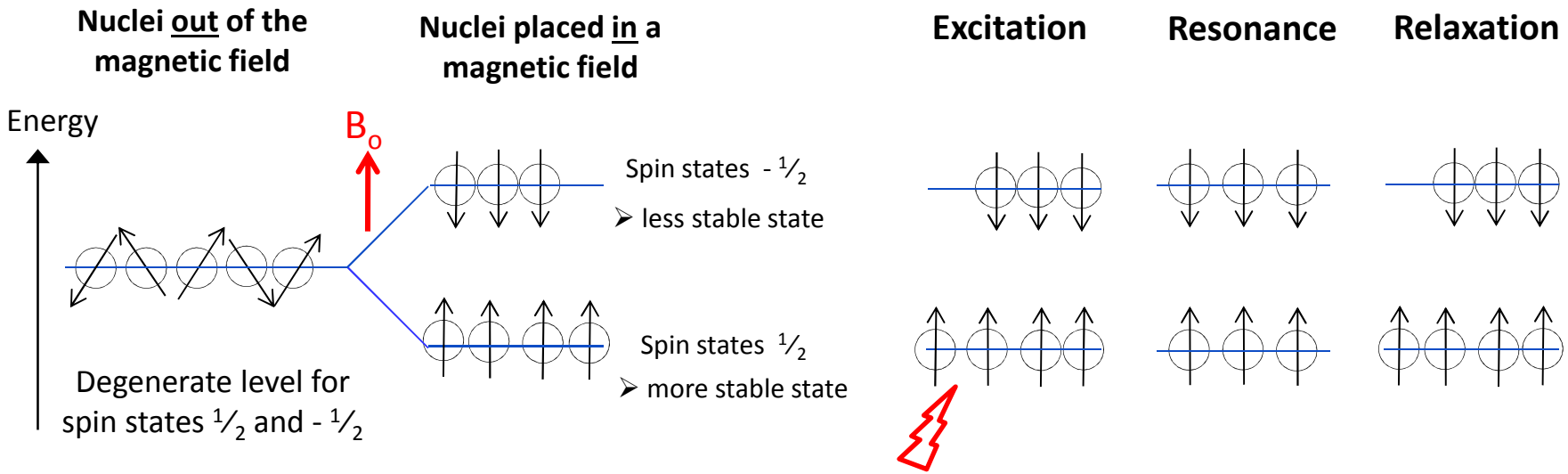
nmr tubes

(dry)
Add (reconstitute in) solvent

MS glass vials

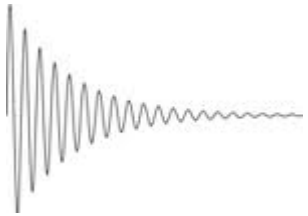
Adapted from "Metabonomics" in the series Methods in Molecular Biology, published by Humana Press, USA –2015, 15-28

NMR, how does it work?



FID NMR signal

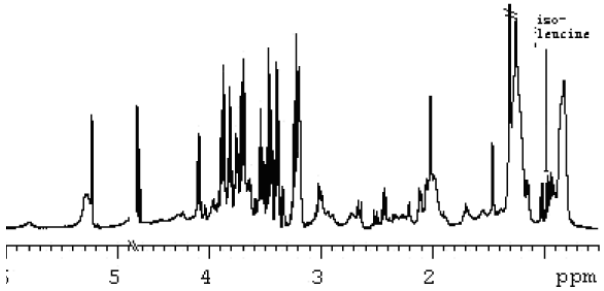
(Free induction decay)



Fourier transform

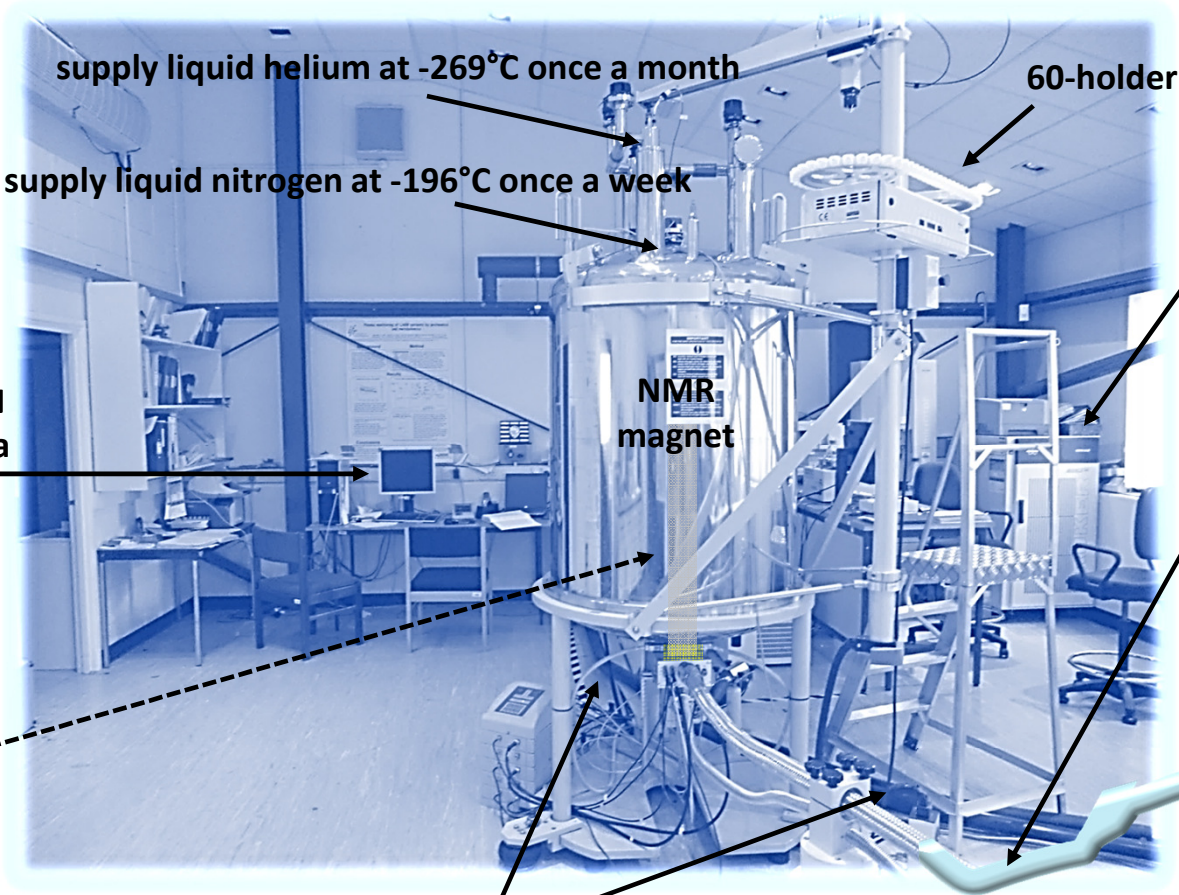
NMR spectrum

(Hz then ppm)



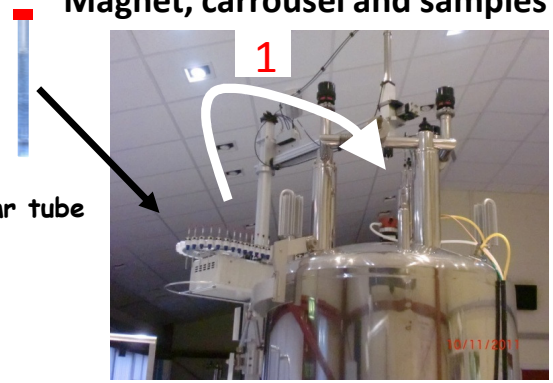
NMR equipment

600 MHz Oxford Instrument NMR spectrometer with cryoprobe and 60-holder autosampler

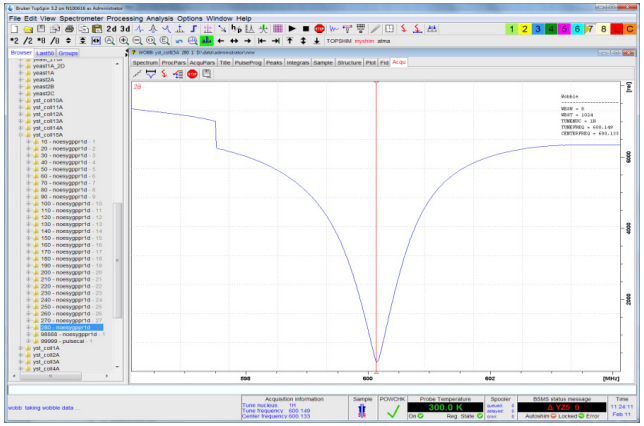


Acquiring an NMR spectrum

Magnet, carrousel and samples in NMR tubes

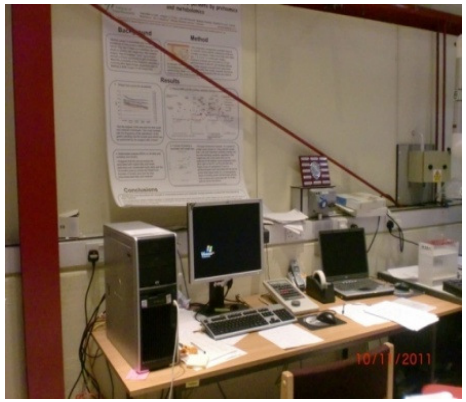


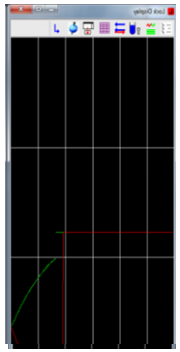
Snapshot of TOPSPIN software



write in the white window the following commands

Workstation (PC with TOPSPIN software)





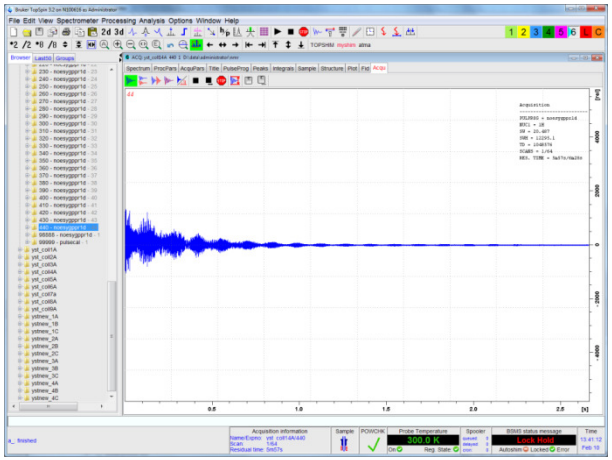
Commands to acquire an ^{1}H NMR spectrum

1. sx 1 (arm moves sample in holder 1 into the magnet)
2. Copy an NMR experiment (set parameters + filename)
3. lock (choose D_2O)
4. atma (tuning)
5. myshim (shimming)
6. zg (acquisition)
7. efp ;apkm (fourier transform, phasing)
8. wpar experiment name
9. iconnmr

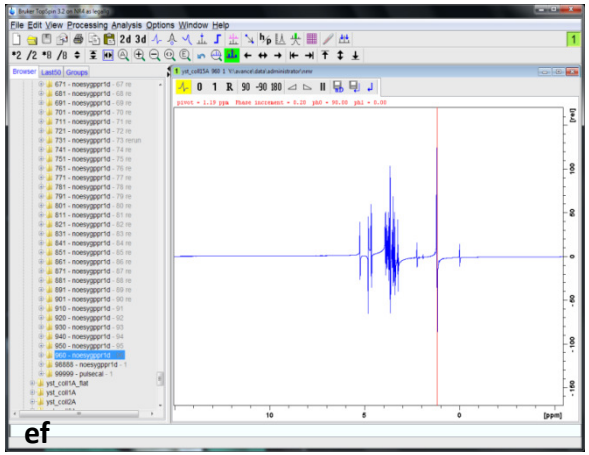
Automation. List of samples can be updated during the run= continuum (~100 samples/24h)

Processing of an NMR spectrum

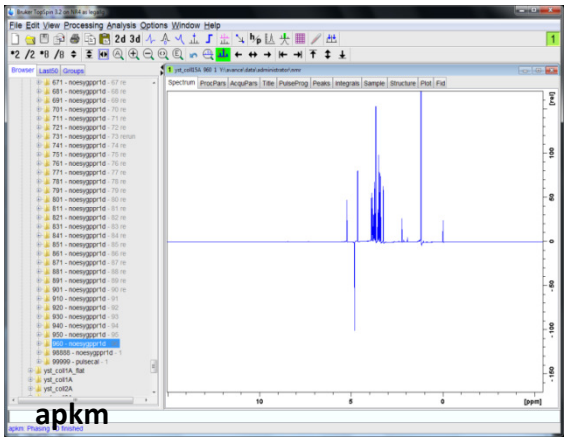
Free induction decay (FID)



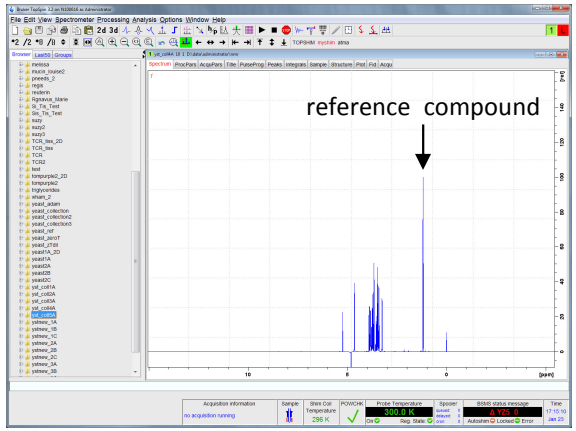
NMR spectrum after Fourier Transform (FT): not phased



NMR spectrum after FT and phasing



NMR spectrum after FT, phasing, baseline correction and setting the reference to zero (final version)

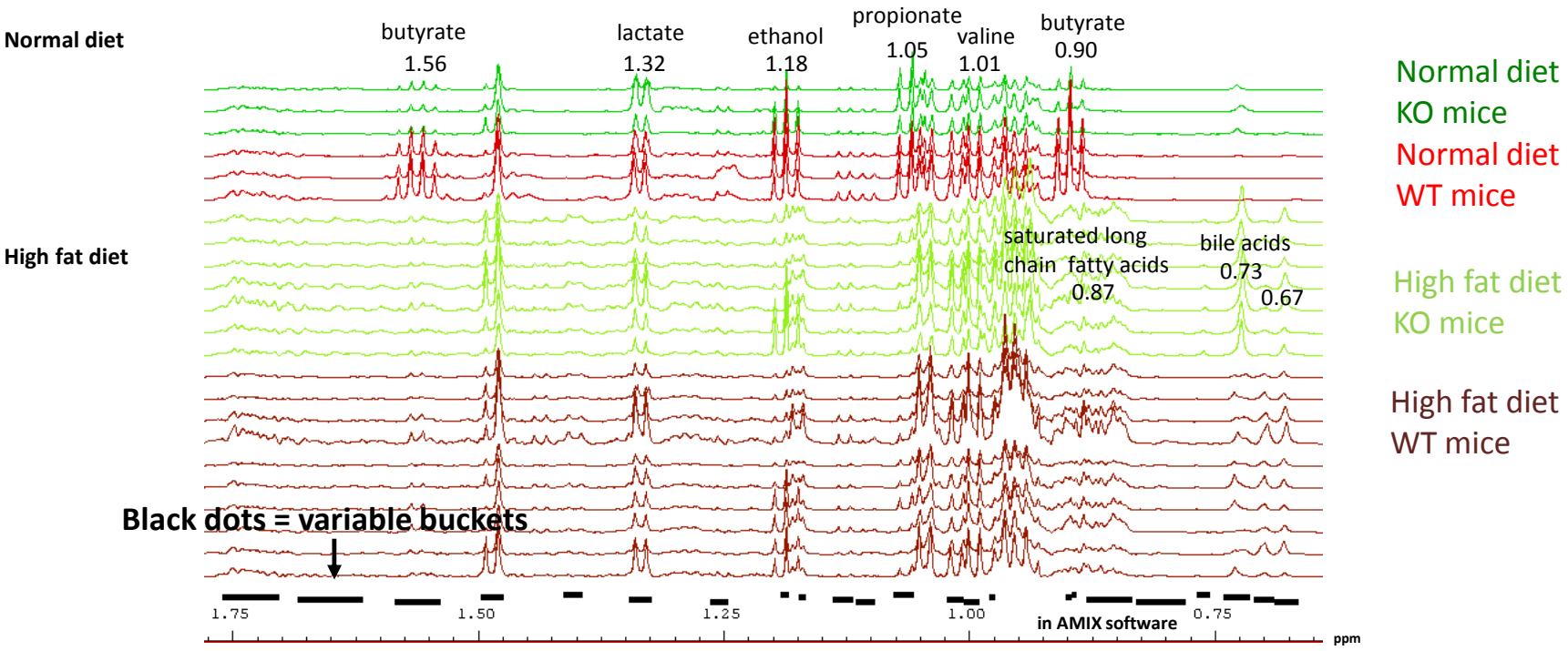


Data processing for NMR

bucketing

ZOOM

^{1}H NMR spectra of mouse faecal pellet samples



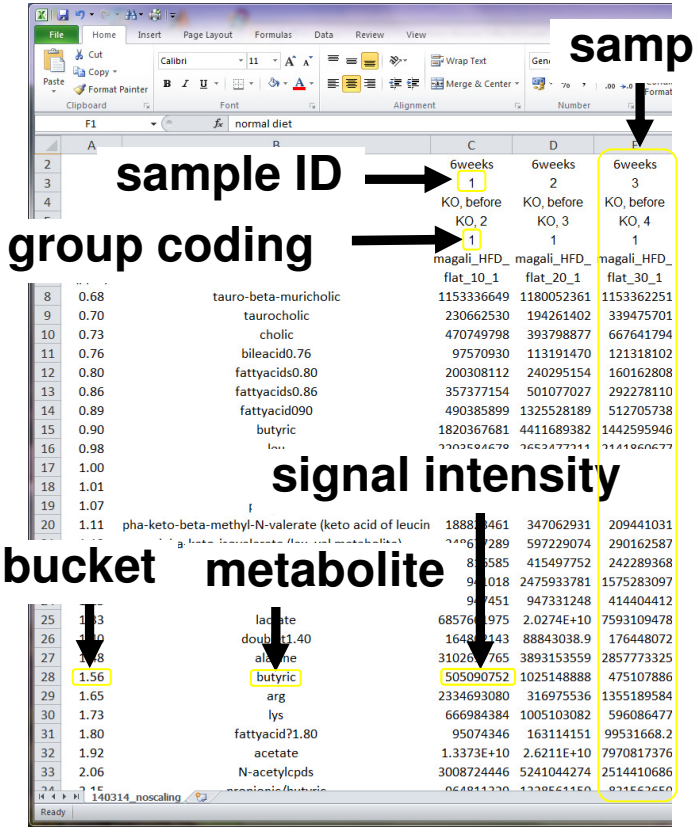
Metabolic changes associated with high fat diet:

- **decrease of the levels of butyrate and propionate**
- **Signals from saturated lipids and bile acids appear**

Data analysis

multivariate analysis

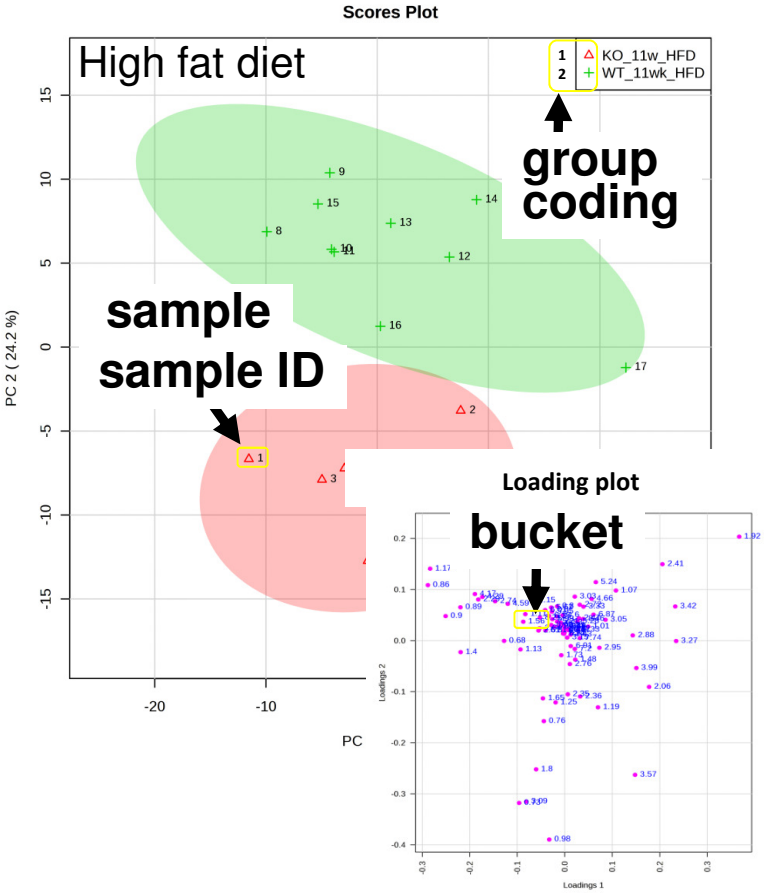
Raw Data



calculation

Principal Component Analysis (PCA)

2 groups - KO vs WT mice

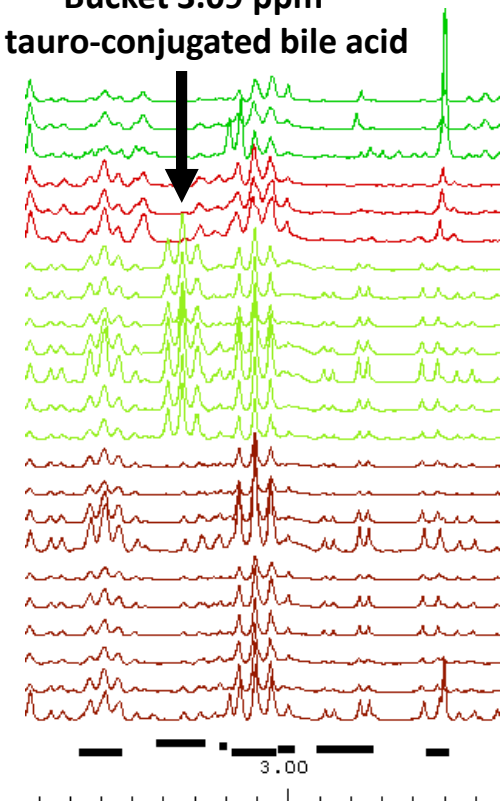


Data analysis

univariate analysis

^{1}H NMR spectra of mouse faecal extracts

Bucket 3.09 ppm – tauro-conjugated bile acid



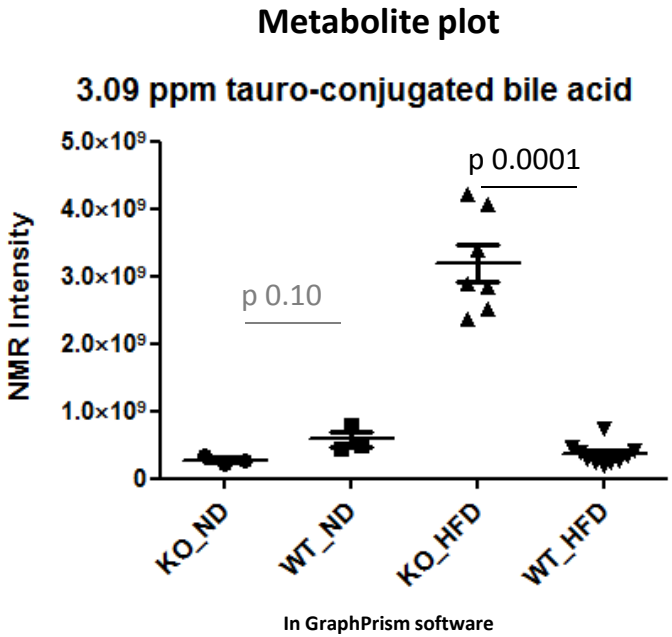
Normal diet KO mice

Normal diet WT mice

High fat diet KO mice

High fat diet WT mice

univariate test

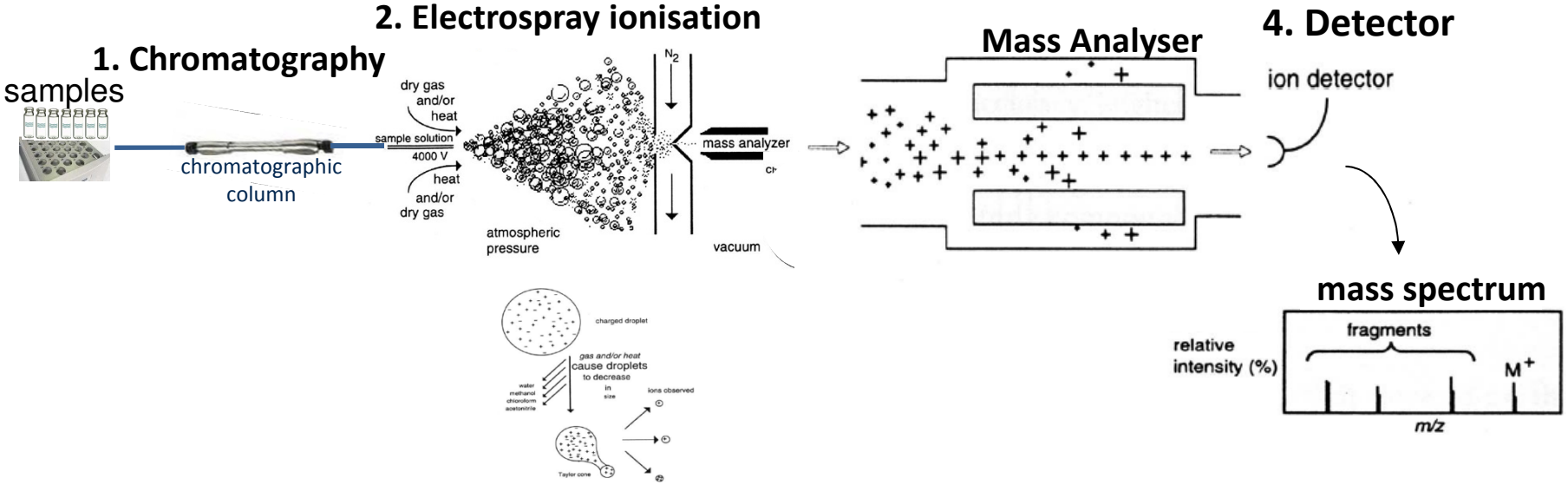


In GraphPrism software

Principle of LC/MS

"Mass spectrometry has been described as the smallest scale in the world, not because of the mass spectrometer's size but because of the size of what it weighs"

Gary Siuzdak- Head of the Scripps centre for metabolomics and mass spectrometry at La Jolla USA



2. ionisation

Molecules turn into ions (M+ and fragments)

3. Mass Analyser

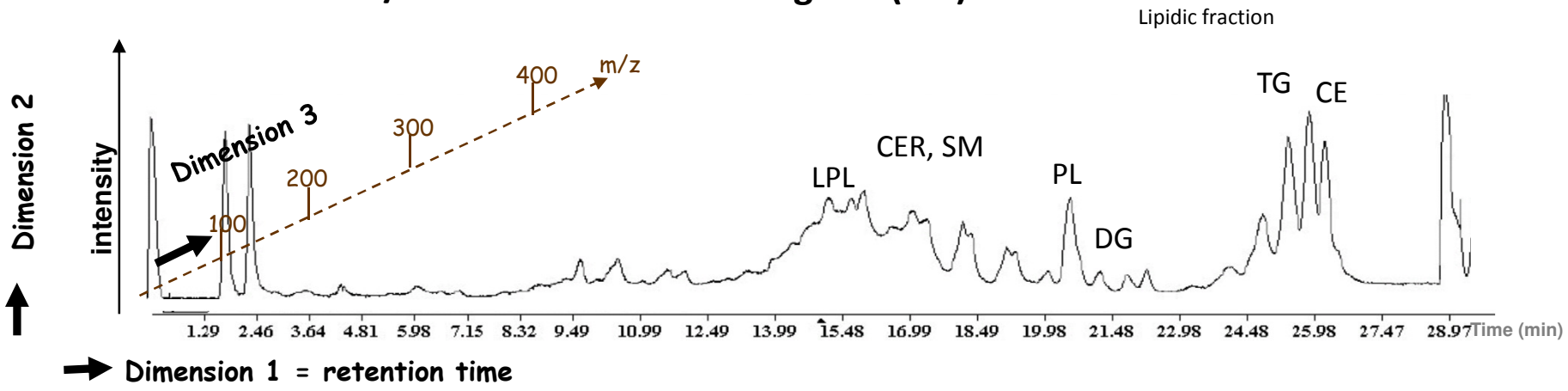
Ions sorted according to their mass over charge ratio (m/z)

NB: multiple chromatographic columns, ionisation methods, analysers (i.e. mass spectrometers, see slide 5), and detectors

Adapted from Mass Spectrometry for Biotechnology, Gary Siuzdak , Academic Press 1996

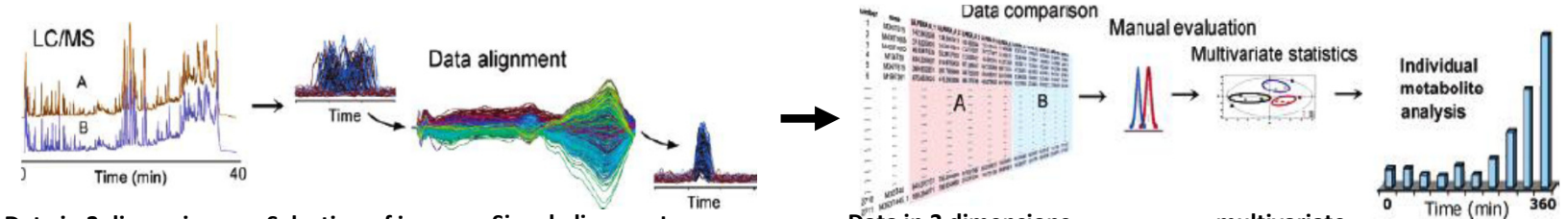
Recording LC/MS data

LC/MS Total Ion Chromatogram (TIC) of mouse liver



Data processing: from 3 to 2 dimensions

XCMS free download software



Data in 3 dimensions — **INPUT**

Selection of ions

Signal alignment

Data in 3 dimensions — **OUPUT**

multivariate analysis

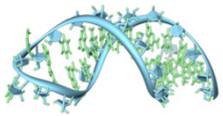
univariate analysis

Want et al, J proteome research, 2007, 6, 459-468

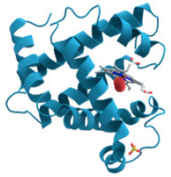
(contrary to transcriptomics and proteomics)

Metabolite identification relies on the analyst's expertise

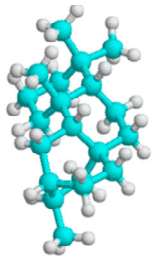
RNA

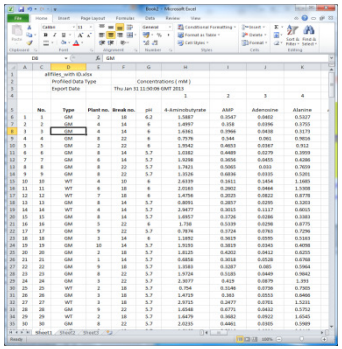


protein



metabolite





Process dataset

Automatic

Search algorithms match experimental sequences to sequences from a pre-defined database

Not automatic

Search spectral databases, build in-house metabolite library, use literature, etc.,

1. Type of chemical entity **2. Record data and processing** **3. Identification**

Metabolite identification

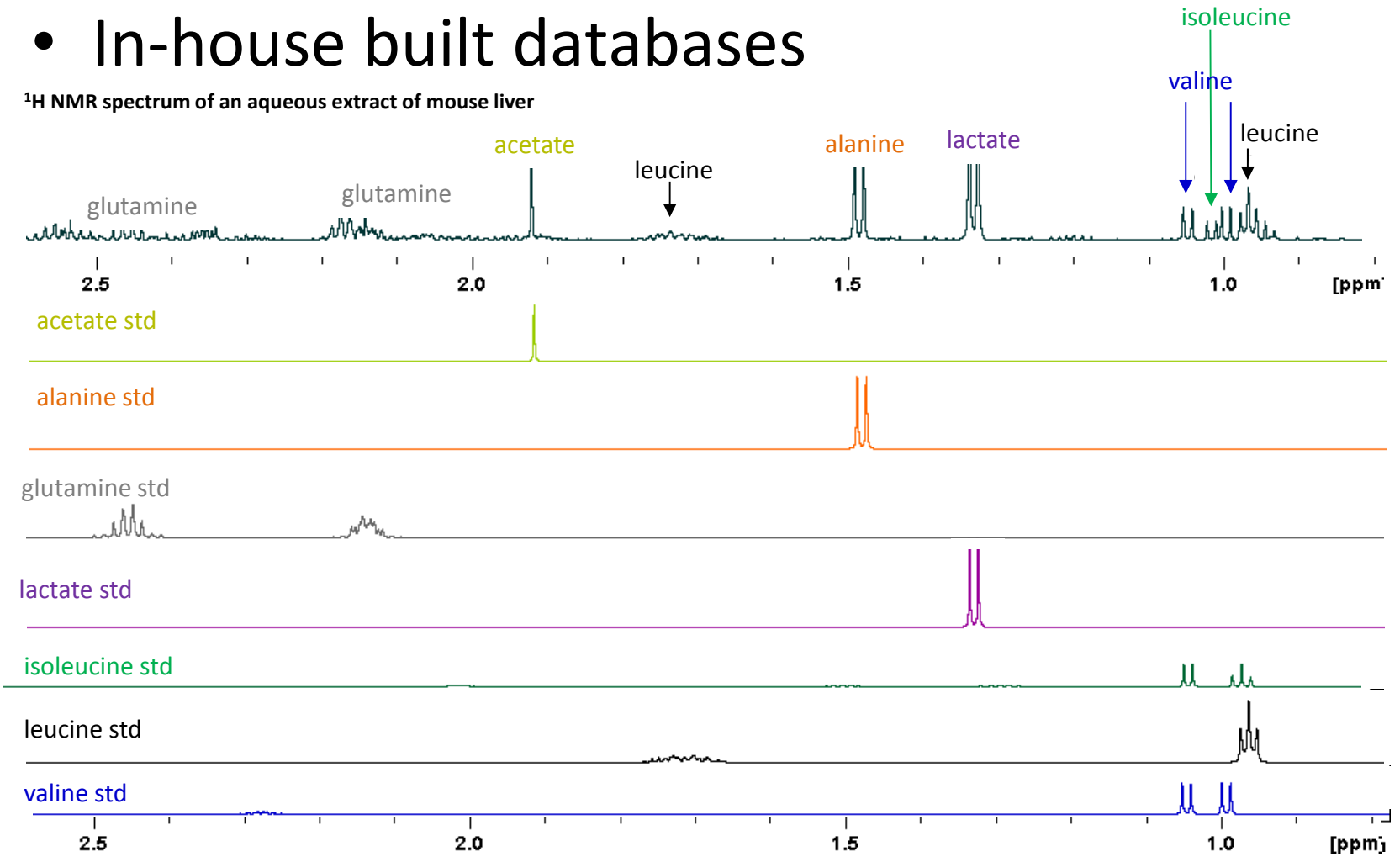
- In-house built databases
- Literature
- 2D NMR spectra or MS/MS analysis followed by consultation of web-based databases

NMR
- Human metabolome database (HMDB) http://www.hmdb.ca/
- Spectral database for Organic Compounds (SDBS) http://sdbs.db.aist.go.jp/sdbs/cgi-bin/cre_index.cgi

MS
- MZedDB the Aberystwyth University High Resolution Mass Spectrometry Laboratory database http://maltese.dbs.aber.ac.uk:8888/hrmet/search/addsearch0.php
- METLIN MS/MS metabolite database https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage#

Metabolite identification

- In-house built databases



Metabolite identification

- Literature metabolites indexed in a table

metabolite identified **^{1}H chemical shift** **^{13}C chemical shift**

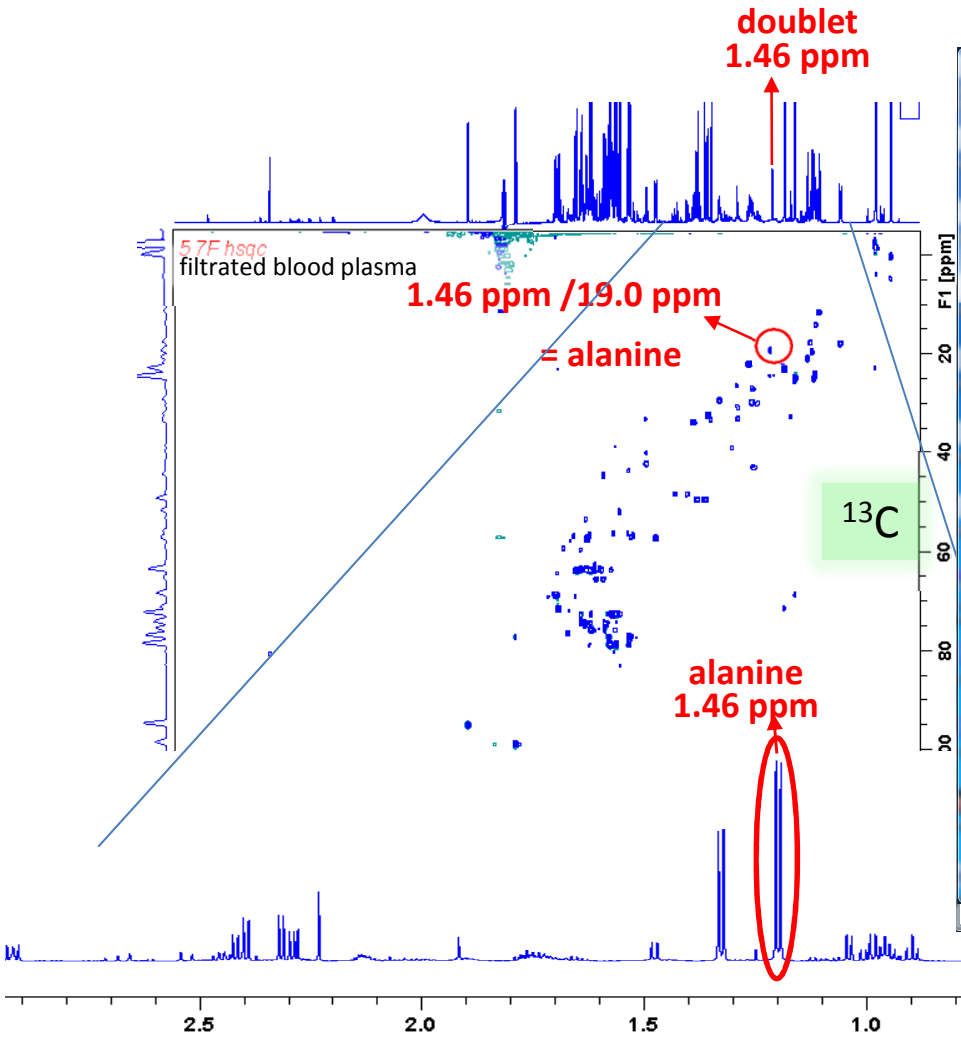
^{1}H and ^{13}C chemical shifts of metabolites identified by 2D NMR in faecal samples

	metabolite	δ^{1}H (ppm)	δ^{13}C (ppm)[a]
1	*n*-butyrate	0.90(t), 1.56(m), 2.16(t)	16.10, 22.14, 42.36
2	propionate	1.06(t), 2.19 (q)	12.90, 33.49
3	valine	0.99(d), 1.05(d), 2.29(m), 3.62(d)	19.44, 20.68, 32.05, 63.50
4	leucine	0.96(d), 0.97(d), 1.70(m), 1.72(m), 3.74(m)	23.68, 24.90, 26.86, 42.44, 56.28
5	isoleucine	0.94(t), 1.02(d), 1.27(m), 1.48(m), 1.99(m), 3.68(d)	13.81, 17.48, 27.49, 27.49, 38.82, 62.40
6	threonine	1.33(d), 3.59(d), 4.26(m)	22.14, 63.34, 68.75
7	isobutyrate	1.07(d), 2.39(m)	22.26, 40.06
8	isovalerate	0.91(d), 1.96(m), 2.06(d)	24.83, 28.75, 50.20
9	*n*-valerate	0.89(t), 1.31(m), 1.53(m), 2.18(t)	16.0, 24.80, 30.92, 40.30
10	*n*-caproate	0.87(t), 1.29(m), 1.31(m), 1.55(m), 2.18(t)	16.0, 33.90, 24.80, 28.49, 40.30
11	*n*-heptanoate	1.31(m)	31.20
12	alanine	1.48(d), 3.78(q)	19.03, 53.62
13	lysine	1.48(m), 1.73(m), 1.91(m), 3.03(t), 3.77(t)	24.38, 29.26, 32.77, 41.98, 57.37
14	arginine	1.70(m), 1.92(m), 3.26(t), 3.77(t)	26.86, 30.45, 43.25, 57.37
15	acetate	1.92(s)	26.10
16	glutamate	2.10(m), 2.36(m), 3.78(dd)	29.81, 36.37, 57.60
17	aspartate	2.69(dd), 2.82(dd), 3.91(dd)	39.36, 39.36, 55.10
18	glycine	3.57(s)	44.34
19	3-phenylpropionate	2.50(t), 2.89(t), 7.27(t), 7.32(d), 7.37(t)	41.96, 34.84, 129.10, 131.20, 131.50
20	3-(4′-hydroxyphenyl)propionate	2.45(t), 2.82(t), 6.85(d), 7.19(d)	42.40, 33.95, 118.22, 132.51
21	tyrosine	3.06(dd), 3.21(dd), 3.95(dd), 6.91(d), 7.20(d)	38.31, 38.31, 58.95, 118.69, 133.66
22	phenylalanine	3.13(dd), 3,29(dd), 4.00(dd), 7.34(m), 7.38(m), 7.44(m)	39.27, 39.27, 58.95, 132.2, 130.5, 131.9
23	tryptophan	3.31(dd), 3.50(dd), 4.07(dd), 7.21(t), 7.28(t), 7.33(s), 7.55(d), 7.74(d)	29.25, 29.25, 58.02, 122.3, 125.0, 127.9, 114.8, 121.4

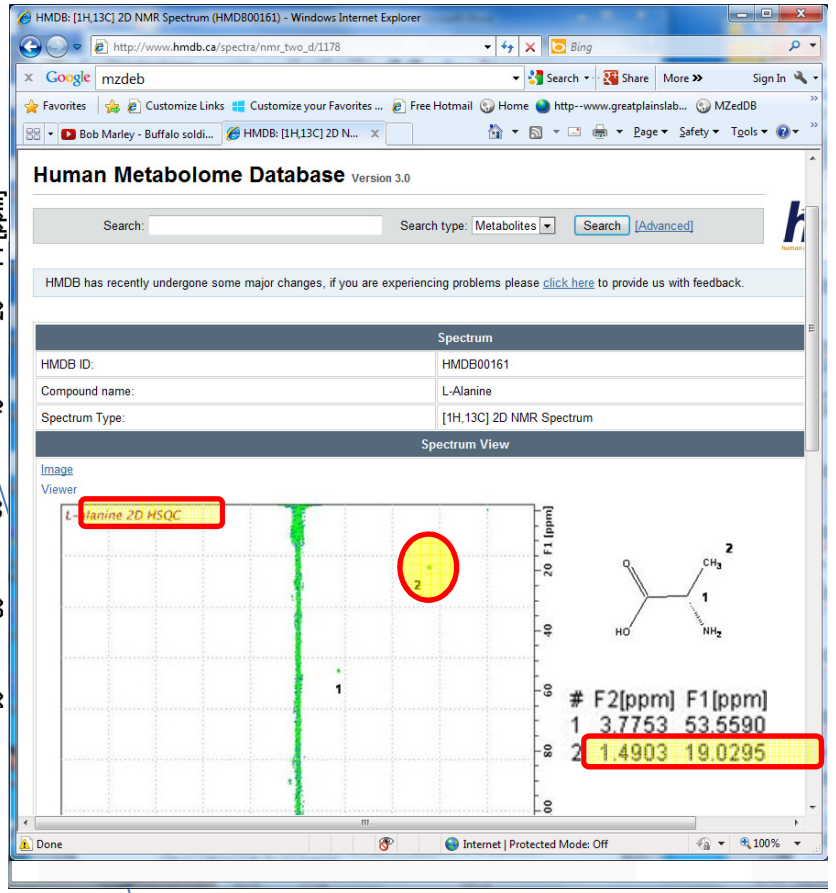
Le Gall et al, Metabolomics of Fecal Extracts Detects Altered Metabolic Activity of Gut Microbiota in Ulcerative Colitis and Irritable Bowel Syndrome, J proteome Res 2011, 10, 4208-18

Metabolite identification

Experimental : 2D NMR HSQC spectrum of a selected sample



Click on L-alanine



List of metabolites

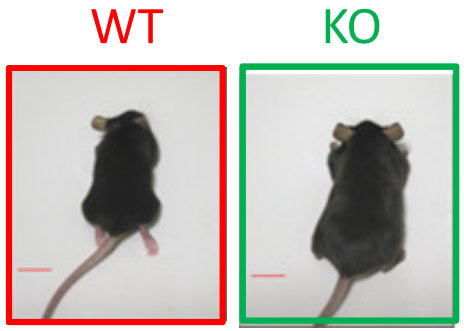
Metabolites detected in ^{1}H NMR spectra of faecal extracts

index	Compound	index (ppm)	type/function	chemical class	multiplicity	average (mmol/kg)
1	**Butyrate**	0.89	product	fatty acid	t	**2.23**
2	**Acetate**	1.9	product	fatty acid	s	**110.35**
3	**Propionate**	1.05	product	fatty acid	t	**10.30**
4	**Isobutyrate**	1.05	product	fatty acid	d	**1.10**
5	**Isocaproate**	0.87	product	fatty acid	d	**0.14**
6	**Isovalerate**	0.9	product	fatty acid	d	**0.40**
7	**Fatty acids C6 and over**	0.85	product	fatty acid	t	**2.57**
8	**Lactate**	3.1	product	fatty acid	q	**15.59**
9	**Lactaldehyde ***	1.37	product	aldehyde	d	**3.98**
10	**Acetaldehyde**	9.67	product	aldehyde	q	**0.07**
11	**Propylene glycol**	1.13	product	alcohol	d	**9.28**
12	**2,3-Butanediol**	1.13	product	alcohol	dd	**4.62**
13	**Ethanol**	1.17	product	alcohol	t	**90.99**
14	**Methanol**	3.35	product	alcohol	s	**0.64**
15	**Formate**	8.46	product	alcohol	s	**2.53**
16	**2-Hydroxyisovalerate**	0.82	intermediate product of branched amino acid	fatty acid	dd	**0.26**
17	**3-Methyl-2-oxovalerate**	1.09	intermediate product of isoleucine	fatty acid	dd	**0.20**
18	**Indole-3-lactate**	7.5	intermediate product of tryptophan	fatty acid	d	**0.18**
			//			
62	**Succinate**	2.39	substrate	energy related acid	s	**14.50**
63	**Pyruvate**	2.36	substrate	energy related acid	s	**0.72**
64	**Pyruvic acid hydrate**	1.46	substrate	energy related acid	s	**5.35**
65	**Adenine**	8.18	substrate / product	nucleobase,side,tide	s	**1.88**
66	**Adenosine monophosphate**	8.6	substrate / product	nucleobase,side,tide	s	**0.03**
67	**2'-Deoxyadenosine**	6.47	substrate / product	nucleobase,side,tide	dd	**0.18**
68	**2'-Deoxyguanosine**	6.3	substrate / product	nucleobase,side,tide	dd	**0.08**
			//			
78	**Uridine**	5.89	substrate / product	nucleobase,side,tide	dd	**0.04**
79	**Uridine monophosphate ***	5.98	substrate / product	nucleobase,side,tide	dd	**0.04**
80	**Xanthine ***	7.83-7.87	substrate / product	nucleobase,side,tide	s	**1.80**
81	**Nicotinate**	8.93	vitamin B3	pyridine	bs	**0.10**

Key to metabolite class

- end-products
- sugars
- amino acids
- osmolytes
- energy related acid
- nucleobase,side,tide
- pyridine

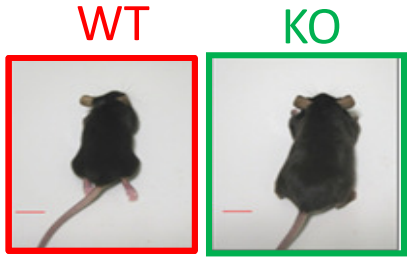
Case Study



metabolite profiling of obese mice

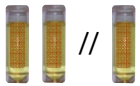
1. Faecal profiling (NMR)

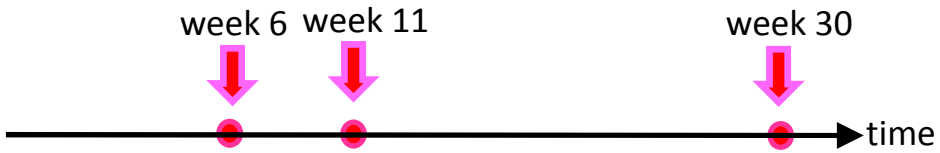
Study design



KO: Knock-out

characterise and compare the metabolic profiles of WT and obese mice (KO)

sample: Faecal pellets 



faeces collection for ^{1}H NMR analysis

Normal diet 

WT KO WT KO WT KO

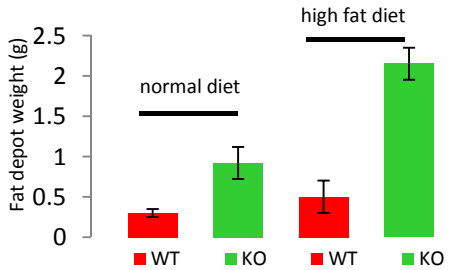
n=16 n=13 n=3 n=3 n=19 n=32

High fat diet 

WT KO

n=10 n=7

Visceral depot
(epididymal)



Phenotype

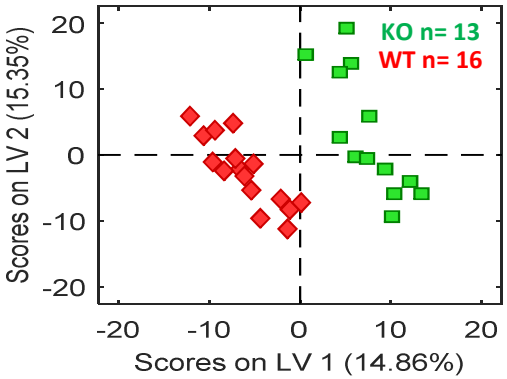
1. Faecal profiling (NMR)

Results

Data analysis - normal diet

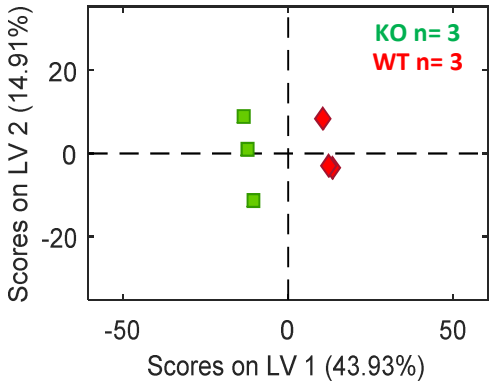
6 weeks old mice

PLS-DA- score plot



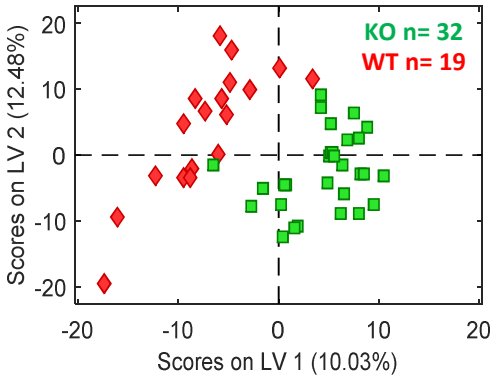
11 weeks old mice

PLS-DA- score plot



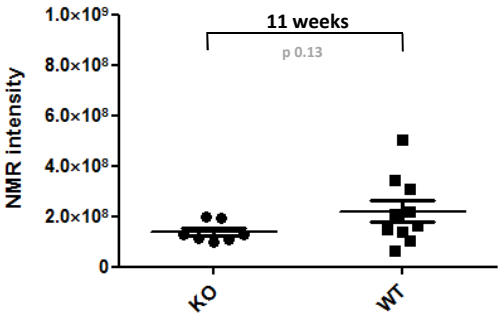
30 weeks old mice

PLS-DA- score plot



High fat diet

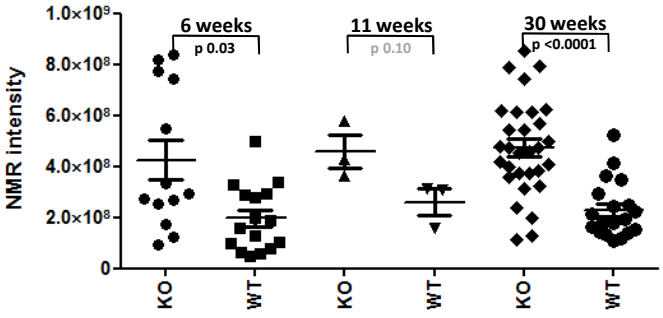
bucket 6.86 ppm



The same phenolic compound has low levels or is absent with high fat diet

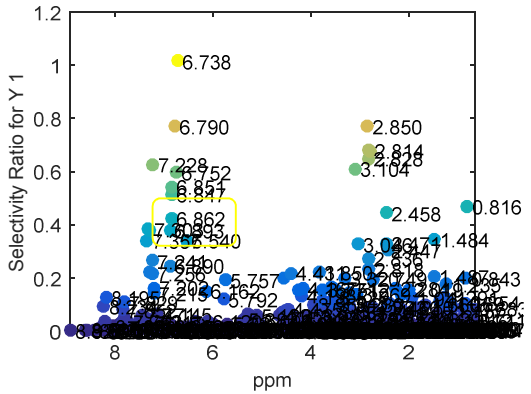
Metabolite plot

bucket 6.86 ppm



In GraphPrism software

PLSDA- marker plot



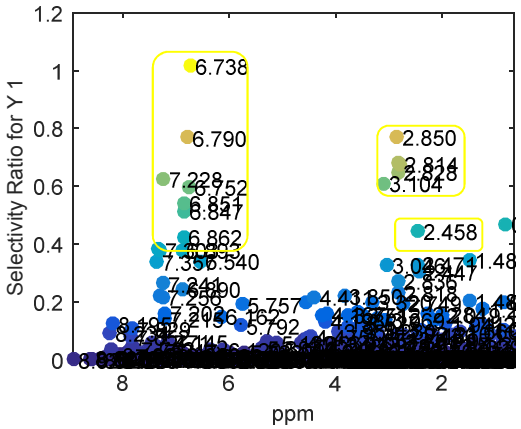
A phenolic compound with higher levels in faeces of KO mice

1. Faecal profiling (NMR)

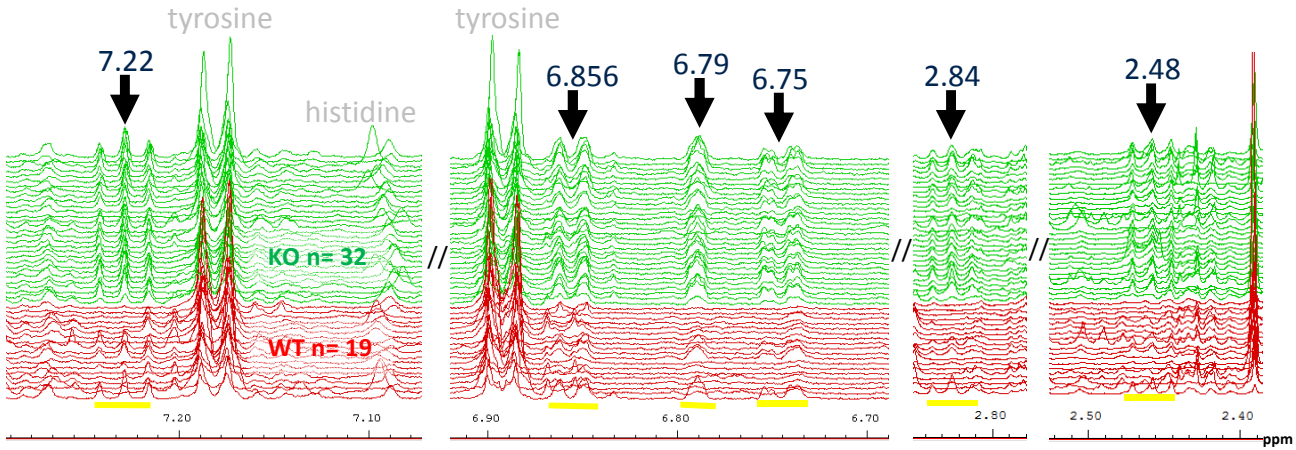
Results

Normal diet, 30 weeks mice

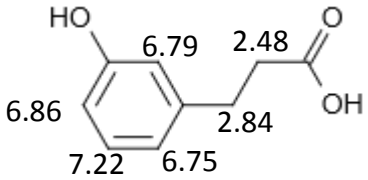
PLSA- marker plot



Raw data: ^{1}H NMR spectra of faecal extracts



3-hydroxyphenylpropionic acid (3HPPA)



A common compound found in faeces of dietary origin (from plants)

Gavaghan et al., Directly coupled high-performance liquid chromatography and Nuclear Magnetic Resonance Spectroscopic with chemometric studies on metabolic variation in Sprague-Dawley rats Anal. Biochem. 2001, 291, 245-252

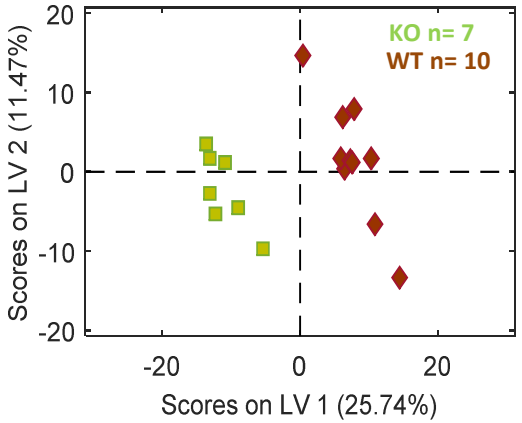
1. Faecal profiling (NMR)

Results

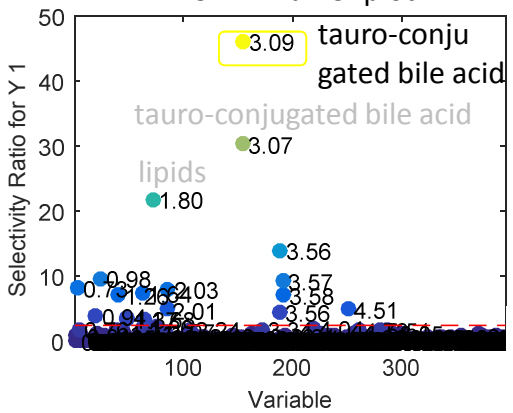
Data analysis - high fat diet

11 weeks old mice

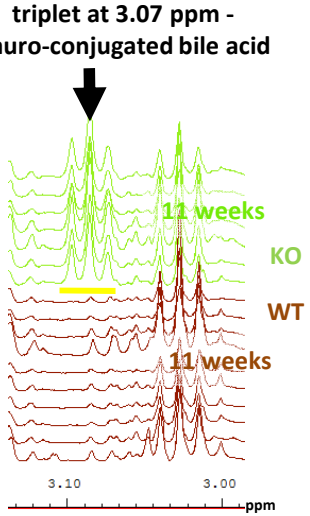
PLS-DA- score plot



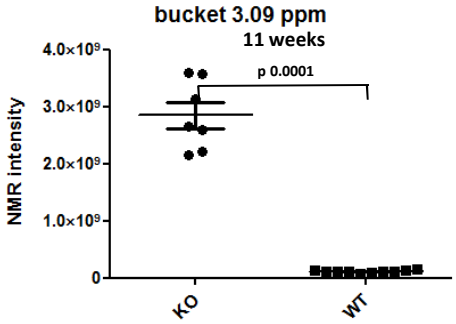
PLSDA- marker plot



Raw data: ^{1}H NMR spectra

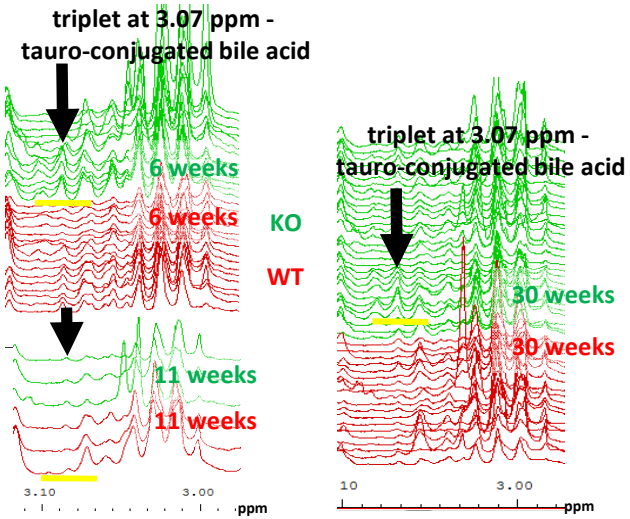


Metabolite plot

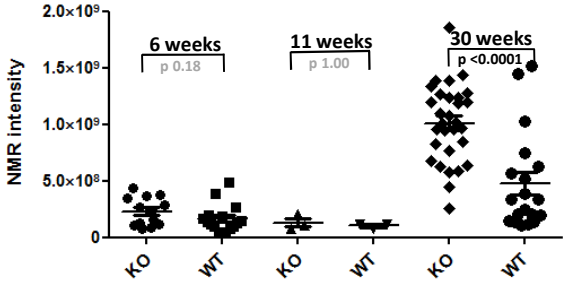


Normal diet

Raw data: ^{1}H NMR spectra



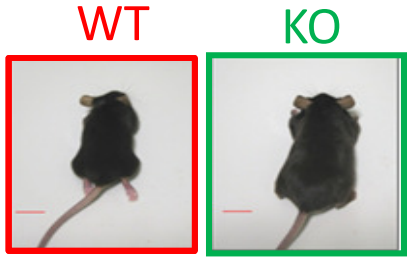
Metabolite plot



A tauro-conjugated bile acid with higher levels in faeces of KO mice at 30 weeks of age or with high fat diet

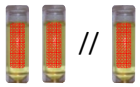
2. Lipidomics (MS)

Study design

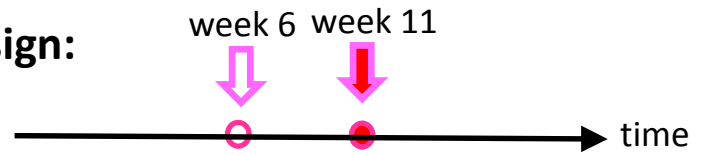


KO: Knock-out

characterise and compare the metabolic profiles of WT and obese mice (KO)

sample: liver extracts 

Study design:



Normal diet 

WT KO

n=5 n=5

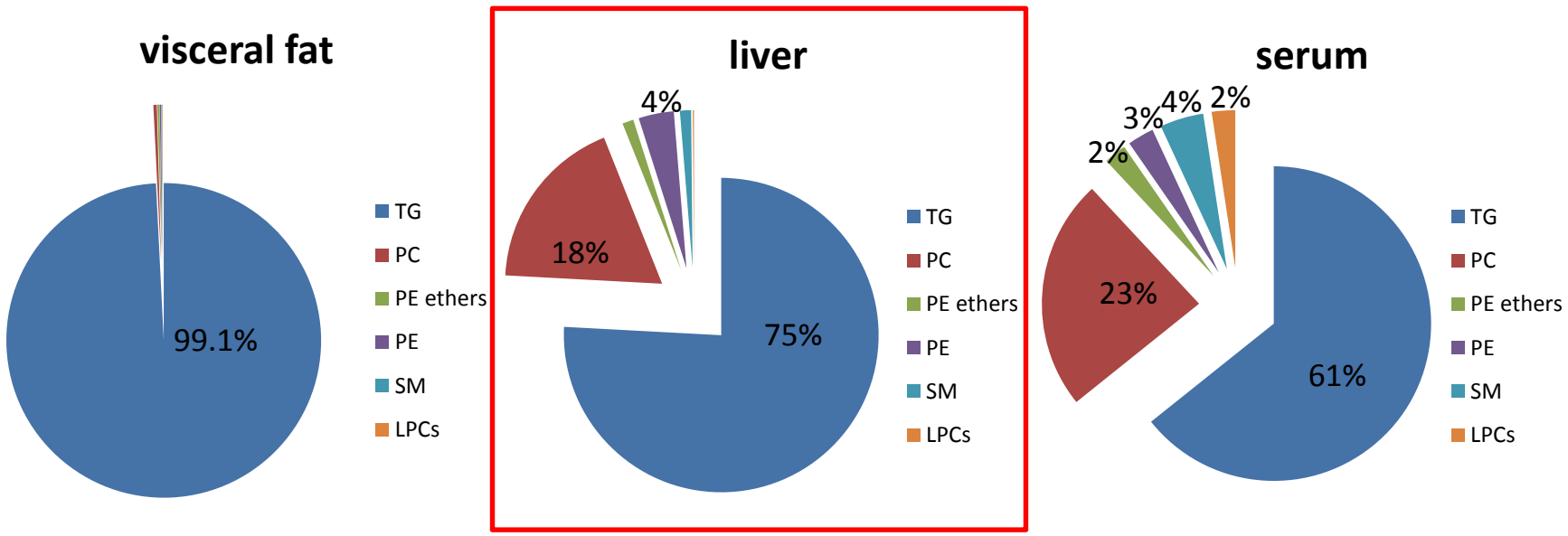
High fat diet 

WT KO

n=5 n=7

Lipid composition in tissues

in humans and mice

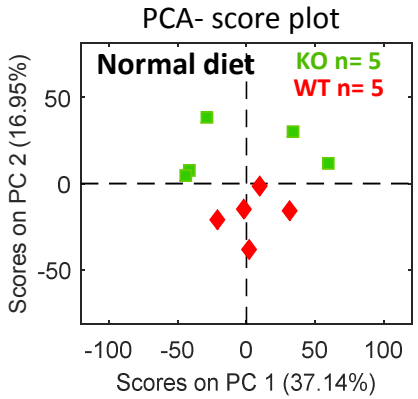


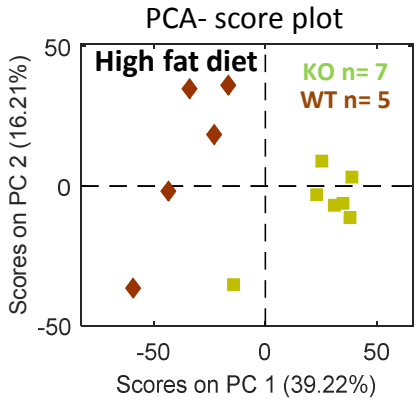
- Adipocytes = Triglycerides
- Liver = Triglycerides; Phospholipids, Phophoethanolamines
- Serum = Triglycerides; Phospholipids, Shingomyelins

2. Lipidomics (MS)

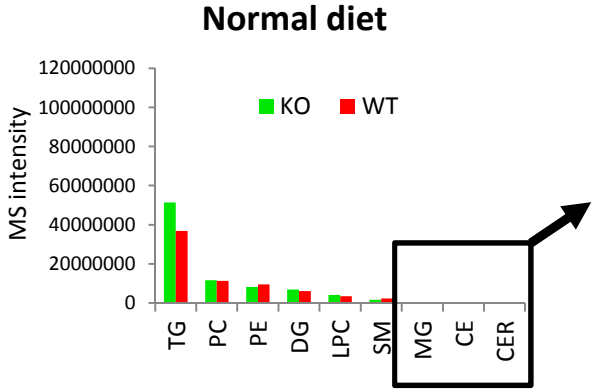
Results

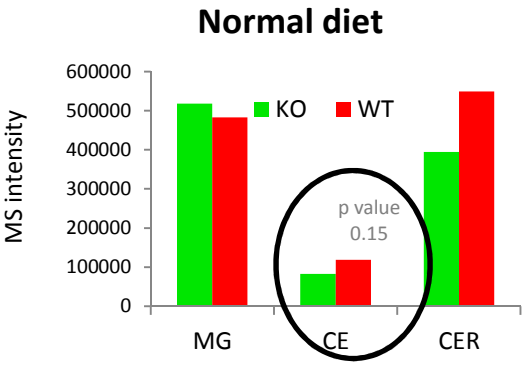
Data analysis

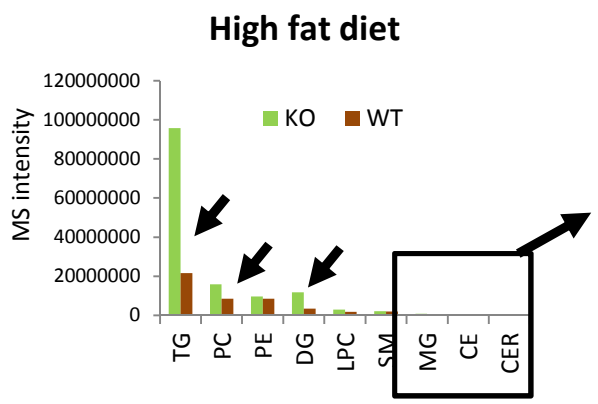


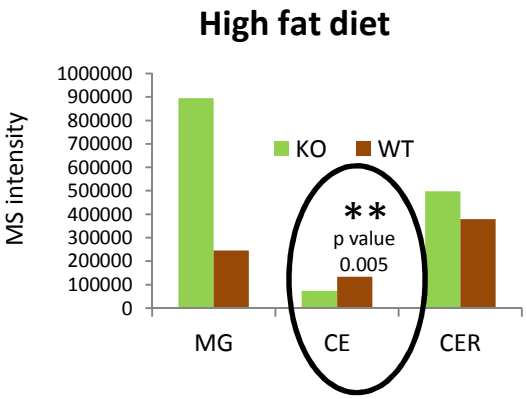


Lipids in the liver





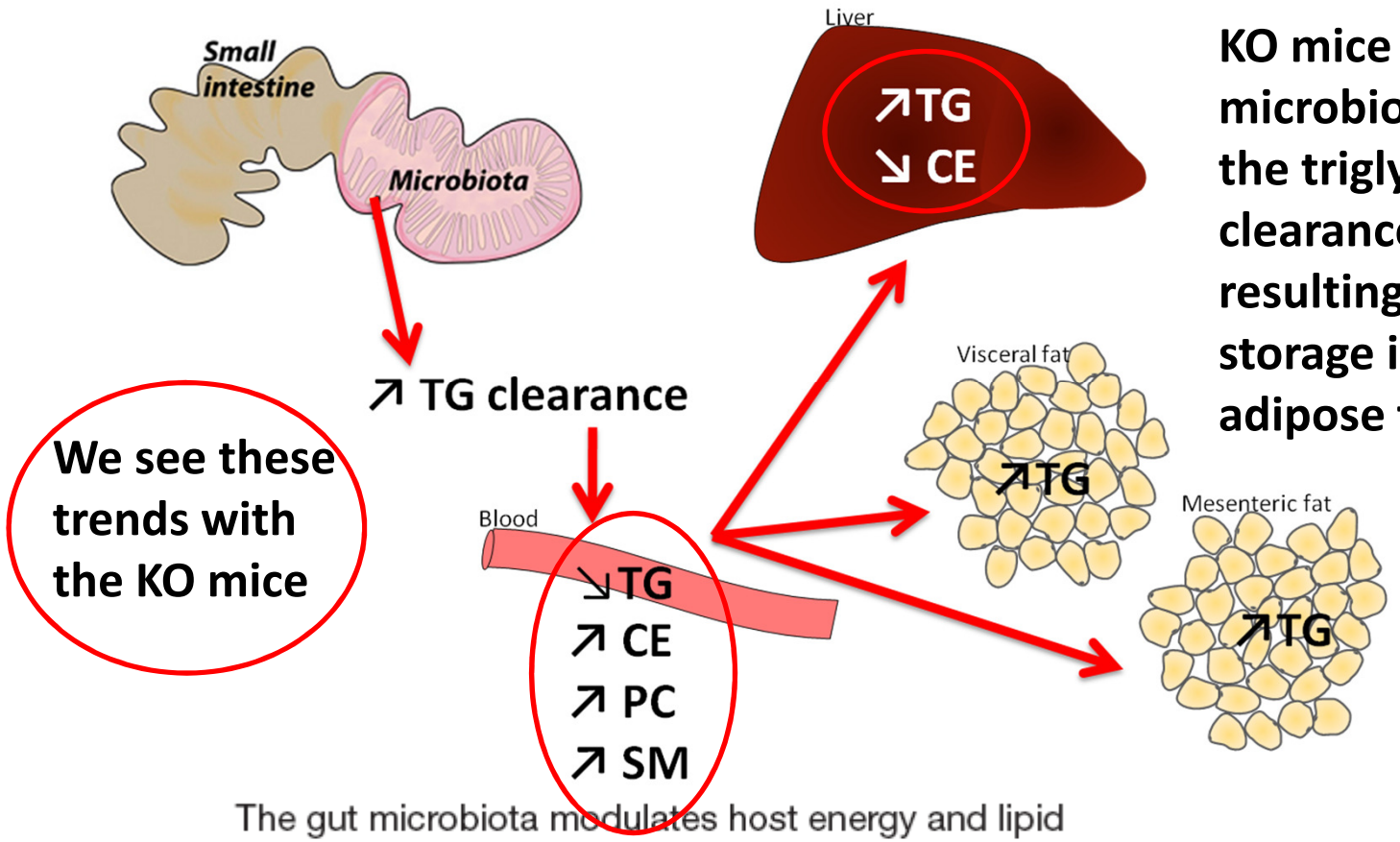




** p < 0.01

- The levels of most lipid species were elevated in KO mice under high fat diet
- Except for cholesterol esters which consistently had lower levels
- Analysis of blood plasma (not shown) revealed a depletion of TGs in KO mice under high fat diet

Proposed model



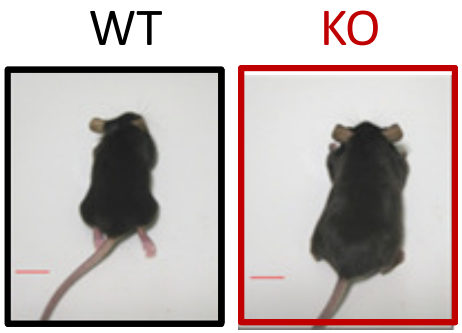
KO mice may have microbiota that increases the triglycerides (TGs) clearance in blood resulting in enhanced TGs storage in the liver and the adipose tissue

The gut microbiota modulates host energy and lipid metabolism in mice[8]

Vidya R. Velagapudi,* Rahil Hezaveh,[†,§] Christopher S. Reigstad,[†,§] Peddinti Gopalacharyulu,* Laxman Yetukuri,* Sama Islam,[†,§] Jenny Felin,[†,§] Rosie Perkins,[†,§] Jan Borén,[†,§] Matej Orešič,*,** and Fredrik Bäckhed[1,†,§]

J Lipid Res. **2010** 51:1101-1

Summary



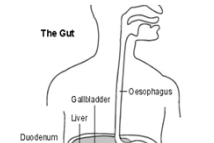
- **The levels of 3-hydroxyphenylpropionic acid (3HPPA) and a tauro-conjugated bile acid were higher in the faeces of KO mice**
- **These results and lipidomic analyses support the hypothesis that the microbiota is involved in generating the phenotype of KO mice (obesity)**
- **High fat diet appears to exacerbate the phenomenon**

Acknowledgements

Simon Carding, Louise Wakenshaw, Magali Sarafian and Ian Colquhoun, Institute of Food Research

Other examples of studies at IFR

Mammalian/Animal Systems



Human interventions

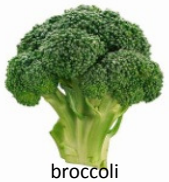

mammalian model


In vitro model

Plants / Food


tomato


Arabidopsis thanalia

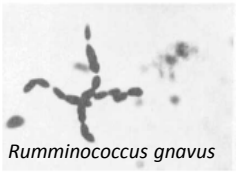

broccoli


green tea

Microbes


Bacteroides thetaiotaomicron


Clostridium botulinum

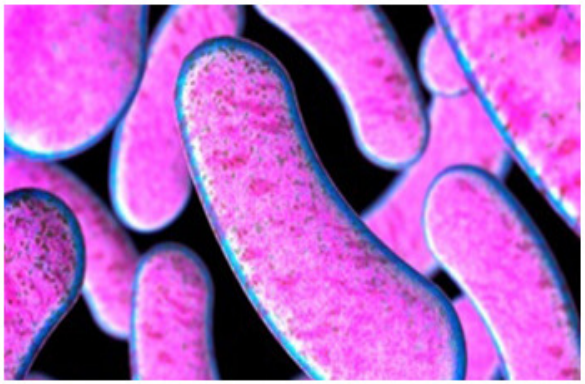

Rumminococcus gnavus

Clostridium botulinum – phylogeny

yeast screening for added-value compounds

Botulinum

Introduction Botulism NHS



Botulism is a very serious infection that is caused by toxins produced by the Clostridium botulinum bacteria

How common is botulism?

Botulism is a rare condition in the UK.

Between 1980 and 2010 there were 33 recorded cases of food-borne botulism in England and Wales. Twenty-six of these were linked to a single outbreak in 1989 that was caused by contaminated hazelnut yoghurt.

Since 1978, there have been 13 cases of infant botulism. None resulted in death.

The number of wound botulism cases has risen sharply in England and Wales over the past 10 years, with 144 cases between 2000 and 2010. This is thought to be due to an increase in people injecting heroin directly into their muscles, a practice often referred to as "skin-popping".

Characteristics

Botulism is a rare but potentially fatal infection caused by toxins produced by Clostridium botulinum bacteria.

The toxins produced by C. botulinum are the most powerful naturally occurring toxins known to science. They attack the nervous system (nerves, brain and spinal cord) and cause paralysis (muscle weakness).

Left untreated, the paralysis will spread to the lungs, causing breathing failure followed by death.

The initial symptoms of botulism include nausea (feeling sick), vomiting and diarrhoea often followed by constipation.

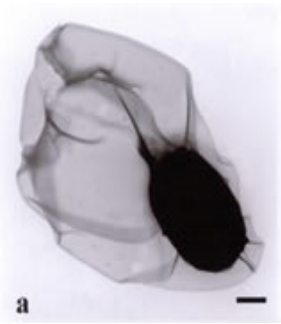
It usually takes 12 to 36 hours after eating contaminated food for more serious neurological symptoms (symptoms affecting the nervous system) to begin. These include double vision, droopy eyelids and slurred speech.

Botulism is a serious condition requiring immediate medical attention. Dial 999 to request an ambulance if you, or someone you know, have the symptoms of botulism.

Read more about the symptoms of botulism.

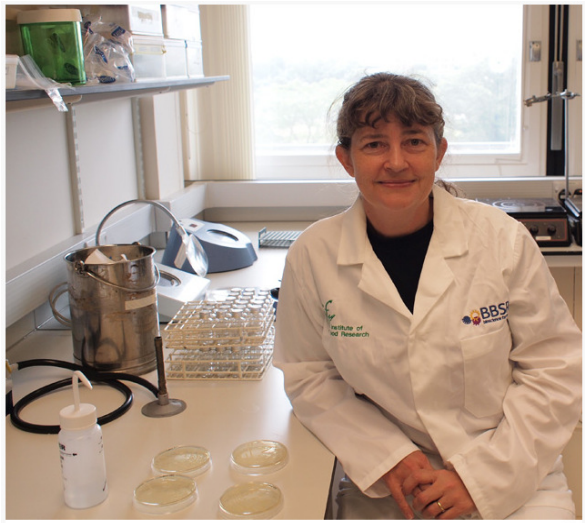
What causes botulism?

Botulism is caused by the bacteria Clostridium botulinum, found in soil, dust and agricultural products such as honey, beans and corn.

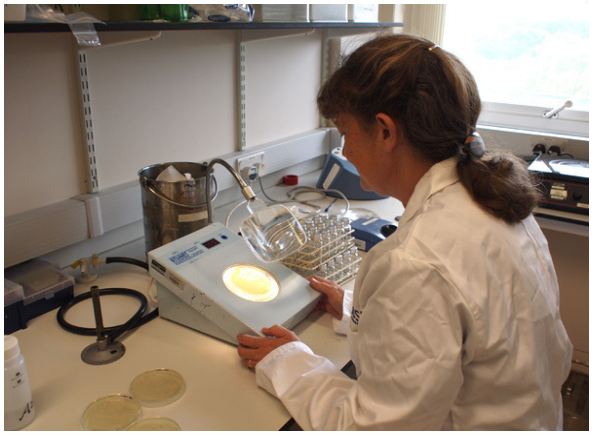


as little as 30 ng of neurotoxin sufficient to cause illness and even death

Clostridium botulinum at IFR



Dr Sandra Stringer



Stringer *et al. BMC Genomics* 2013, **14**:333
http://www.biomedcentral.com/1471-2164/14/333

BMC Genomics

RESEARCH ARTICLE Open Access

Genomic and physiological variability within Group II (non-proteolytic) *Clostridium botulinum*

Sandra C Stringer[1*], Andrew T Carter[1], Martin D Webb[1], Ewelina Wachnicka[1], Lisa C Crossman[2], Mohammed Sebaihia[3,4] and Michael W Peck[1]

Abstract

Background: *Clostridium botulinum* is a group of four physiologically and phylogenetically distinct bacteria that produce botulinum neurotoxin. While studies have characterised variability between strains of Group I (proteolytic) *C. botulinum*, the genetic and physiological variability and relationships between strains within Group II (non-proteolytic) *C. botulinum* are not well understood. In this study the genome of Group II strain *C. botulinum* Eklund 17B (NRP) was sequenced and used to construct a whole genome DNA microarray. This was used in a comparative genomic indexing study to compare the relatedness of 43 strains of Group II *C. botulinum* (14 type B, 24 type E and 5 type F). These results were compared with characteristics determined from physiological tests.

Results: Whole genome indexing showed that strains of Group II *C. botulinum* isolated from a wide variety of environments over more than 75 years clustered together indicating the genetic background of Group II *C. botulinum* is stable. Further analysis showed that strains forming type B or type F toxin are closely related with only toxin cluster genes targets being unique to either type. Strains producing type E toxin formed a separate subset. Carbohydrate fermentation tests supported the observation that type B and F strains form a separate subset to type E strains. All the type F strains and most of type B strains produced acid from amylopectin, amylose and glycogen whereas type E strains did not. However, these two subsets did not differ strongly in minimum growth temperature or maximum NaCl concentration for growth. No relationship was found between tellurite resistance and toxin type despite all the tested type B and type F strains carrying *tehB*, while the sequence was absent or diverged in all type E strains.

Conclusions: Although Group II *C. botulinum* form a tight genetic group, genomic and physiological analysis indicates there are two distinct subsets within this group. All type B strains and type F strains are in one subset and all type E strains in the other.

Aim in paper: classify the strains in Group II *Clostridium botulinum*

1. Classification based on growth on substrates

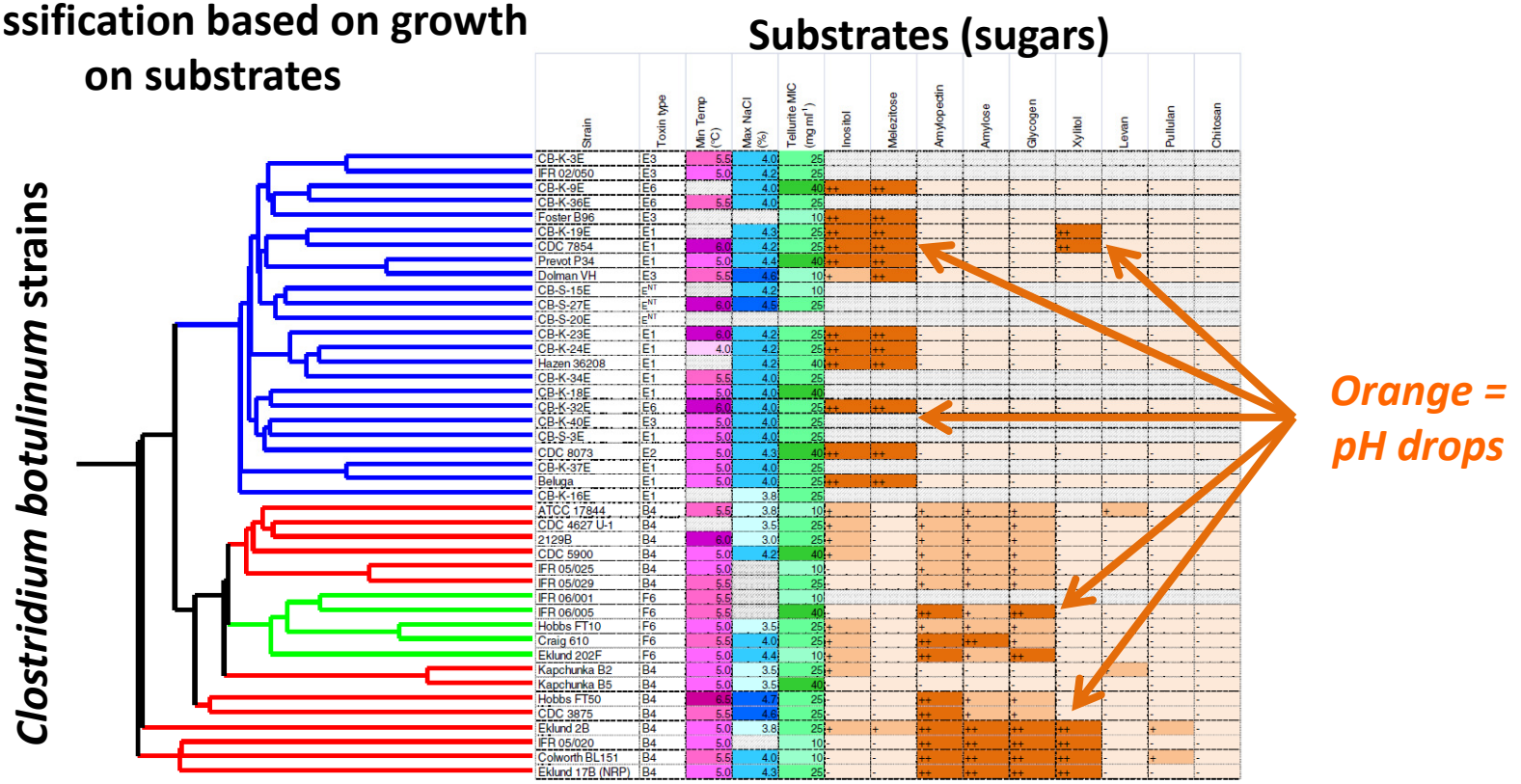


Figure 6 Physiological characteristics of strains of Group II *C. botulinum*. The minimum temperature and maximum NaCl concentration at which growth was observed, the minimum concentration of tellurite required to prevent growth and the ability to ferment selected carbohydrates was tested on strains representing different clades. Acid production was measured in a PY basal medium with 10 g l^{-1} added carbohydrate. A carbohydrate was considered to be fermented if the final pH was more than 0.5 units less than inoculated medium in the absence of carbohydrate. A pH reduction of 0.5-1.0 units was noted as acid production (+) and >1.0 was noted as strong acid production (++).

Stringer *et al. BMC Genomics* 2013, **14**:333
http://www.biomedcentral.com/1471-2164/14/333

Next: identify compounds that classify the strains in Group II *Clostridium botulinum*

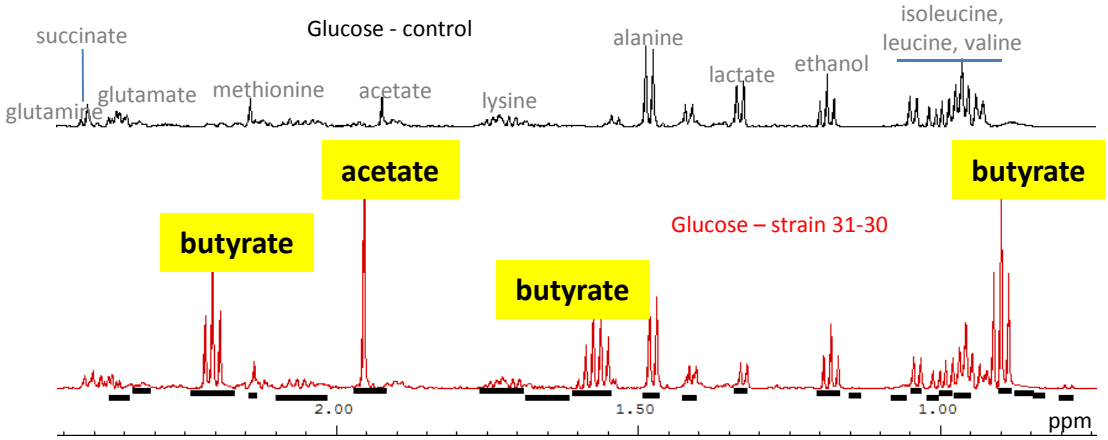
Samples:

- **Strains** type B, E and F
- **Substrates** were the medium alone (control), medium + inositol, melizitose, amylopectin, amylose, glycogen, xylitol, levan, pullulan, chitosan, glucose, ribose and maltose

Results:

- 428 ^{1}H NMR spectra of spent medium were recorded
- Metabolites detected: acetate, butyrate, propionate, formate, fumarate, propanediol, trimethylamine, 5,6-dihydrouracil and maltose

^{1}H NMR SPECTRA



Origin of trimethylamine: Stickland fermentation

(oxidation and reduction of amino acids to organic acids)

amino acid+ 2 betaine + 2 H_20 ⟶ acid + CO_2 + NH_3 + 2 trimethylamine + 2 acetate

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Feb. 1983, p. 474-483

Compounds that classify the strains in Group II *Clostridium botulinum*

The acids responsible for the pH drop were acetate, butyrate, formate and propionate

Substrates (sugars)

Clostridium botulinum strains

1		strains for NMR_Gwen.xls		"Strains"											
2	Strain details			Prepared for NMR	Sugar fermentation test results (final pH of medium)										
4	Clade	Strain no.	Original name	Toxin	Carb./ Strain	No sugar	Inositol	Melezitose	Amylopectin	Amylose	Glycogen	Xylitol	Levan	Pullulan	Chitosan
5	1	81-30	Eklund 17B	B	81-30	5.85	5.85	5.69	**4.74**	5.02	4.76	5.03	6.78	5.88	6.67
6	1	83-01	Eklund 2B	B	83-01	5.95	5.44	5.43	**4.81**	**4.98**	4.87				
7	1	86-17	Colworth 151	B	86-17	6.14	5.84	5.91	**4.79**	**4.88**	4.79	**4.99**	6.74	5.56	6.1
8	1	05-20	2547/1	B	05-20	6	5.95	5.86	**4.69**	**4.86**	**4.8**	5.02	6.88	6.01	6.77
10	2	81-23	Hobbs FT 50	B	81-23	5.99	5.57	5.87	5.06	5.12	5.04	5.9	6.75	5.64	6.63
11	2	87-02	CDC 3875	B	87-02	6.1	5.68	5.94	**5.05**	5.25	5.03	5.93	6.86	5.98	6.66
12	2	87-04	CDC 4672 U-1	B	87-04	5.97	5.51	5.88	5.22	5.27	5.17	5.87	6.77	5.95	6.55
13	2	90-04	Prevot 59	B	90-04	5.96	5.11								
14	2	93-06	CDC 5900	B	93-06	6.08		5.95	5.44	5.36	5.46	6	6.81	6.01	6.69
15	2	93-10	Kapchunka B2	B	93-10	5.86	5.54	5.72	5.84	5.87	5.83	5.78	6.3	5.88	6.68
16	2	93-11	Kapchunka B5	B	93-11	6.13	5.59	5.93	6.02	6.1	6.05	5.99	6.82	5.9	6.53
17	2	02-51	CB-R-25 Bnp	B	02-51	5.9	5.53	5.79	5.17	5.26	5.2	5.77	6.35	5.87	6.7
18	2	05-25	1074/1	B	05-25	5.86	5.75	5.78	5.25	5.43	5.15	5.84	6.69	5.85	6.82
19	2	05-29	1082/1	B	05-29	5.87	5.6	5.74	5.28	5.47	5.29	5.81	6.74	5.83	6.84
21	3	81-26	Beluga	E	81-26	6.12	**5.03**	**4.93**	6.08	6.11	6.13	6.05	6.87	6.09	6.66
22	3	81-27	Foster B56	E	81-27	5.99	**4.73**								
23	3	81-31	Hazen 36208	E	81-31	5.88	4.97	**4.75**	5.83	5.88	5.82	5.76	6.72	5.85	6.52
24	3	86-21	Sebald P34	E	86-21	6.07	**4.94**	**4.67**							
25	3	87-01	DOLMAN VH	E	87-01	5.95	5.46	**4.8**		5.92	5.88	5.8	6.68	5.91	6.56
26	3	93-07	CDC 7854	E	93-07	6.24	**4.71**	**4.94**	6.16	6.25	6.2	**4.96**	6.85	6.2	6.66
27	3	93-08	CDC 8073	E	93-08	5.99	**4.71**	**4.74**	5.97	6	5.98	5.9	6.84	6.01	6.6
28	3	02-07	CB-K-9E	E	02-07	6.16	**4.74**	**4.8**	6.16	6.17	6.09				
29	3	02-10	CB-K-19E	E	02-10	5.99	**4.76**	**4.83**	6	6.03	5.97	**4.84**			
30	3	02-15	CB-K-24E	E	02-15	6.1	**5.07**	**4.84**	6.05	6.04	6.02	5.97	6.76	6	6.58
31	3	02-22	CB-K-32E	E	02-22	5.92	4.97	**4.63**	5.87	5.92	5.89	5.8	6.75	5.92	6.62
32	3	02-43	CB-S-20E	E	02-43	6.07	**4.89**	**4.7**	6.05	6.14	6.13	6.09	6.82	6.14	6.64
34	2	86-32	Eklund 202F	F	86-32	5.92	5.24	5.79	**4.93**	5.15	**4.9**	5.83	6.81	5.91	6.6
35	2	86-33	Hobbs FT 10	F	86-33	5.97	5.43	5.8	5.14	5.14	5.1	5.8	6.77	5.85	6.56
36	2	86-34	Craig 610	F	86-34	6.01	5.44	5.89	**5**	**4.99**	5.04	5.86	6.71	6.03	6.63
37	2	06-05	4392-5	F	06-01	5.99	5.7	5.86	**4.98**	5.11	**4.97**	5.83	6.75	5.98	6.62
39		Contro	Uninoculated	No	Control (no inoculum)	6.58	6.57	6.6	6.61	6.65	6.53	6.59	7.48	6.59	7.07

Orange = pH drops

acetate and butyrate

acetate, butyrate and propionate

acetate and propionate

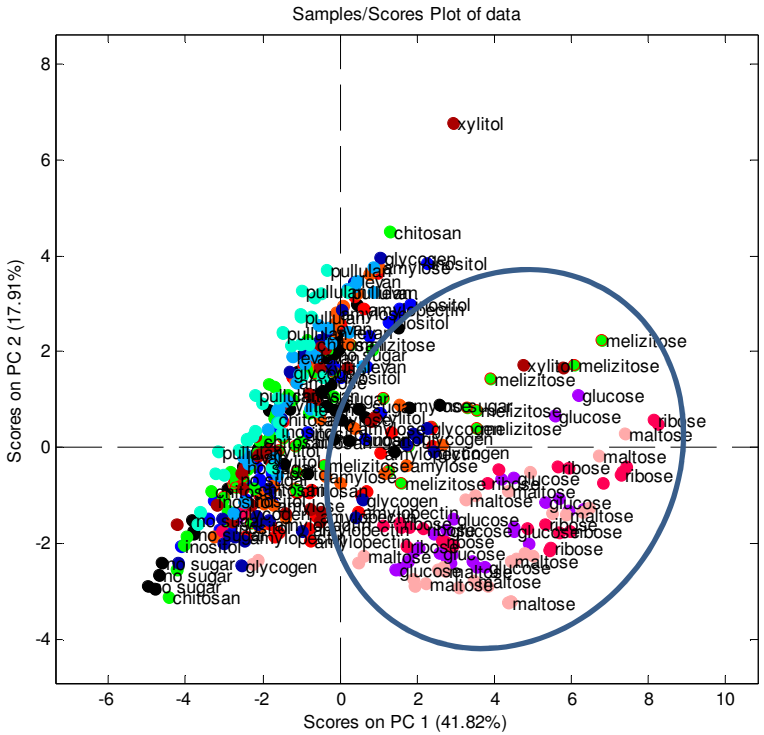
acetate and butyrate

butyrate and formate

Multivariate analysis on ^{1}H NMR data

PCA on 428 ^{1}H NMR spectra

Coloured coded according to substrates



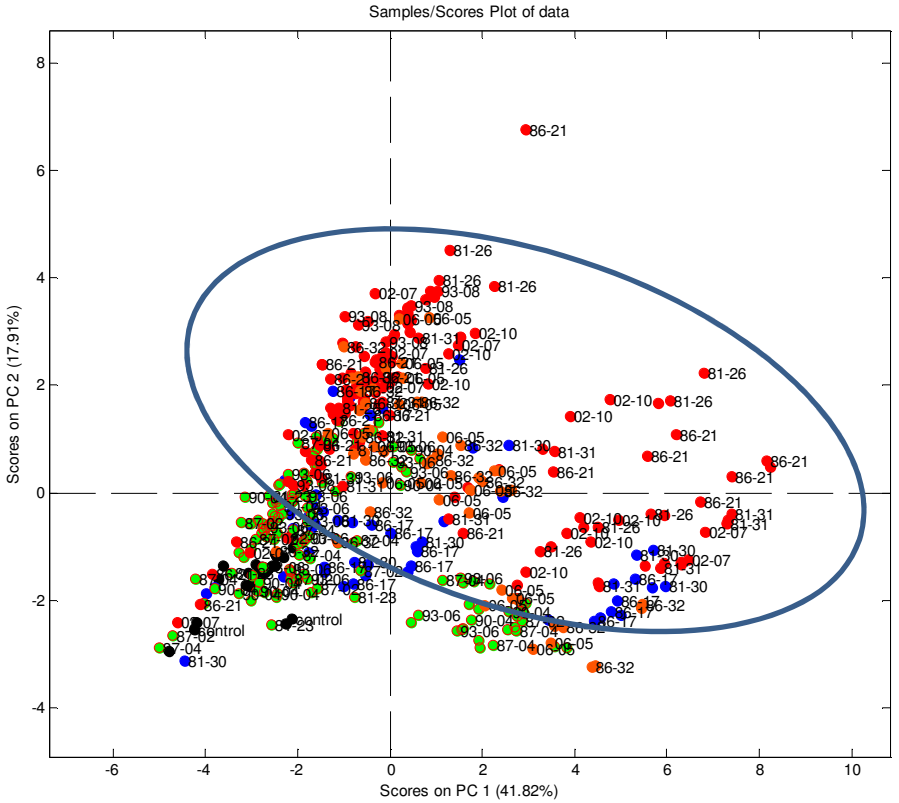
PC1 reflects an increase fermentation (butyrate and acetate) on glucose, ribose, maltose

Multivariate analysis on ^{1}H NMR data

PCA on 428 ^{1}H NMR spectra

Coloured coded according to clades

- Clade 1
- Clade 2
- Clade 3
- Clade 4

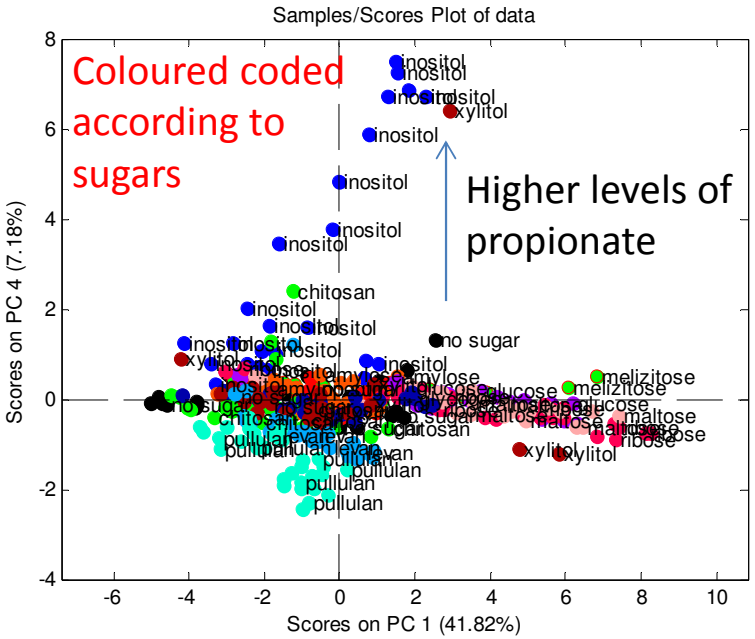


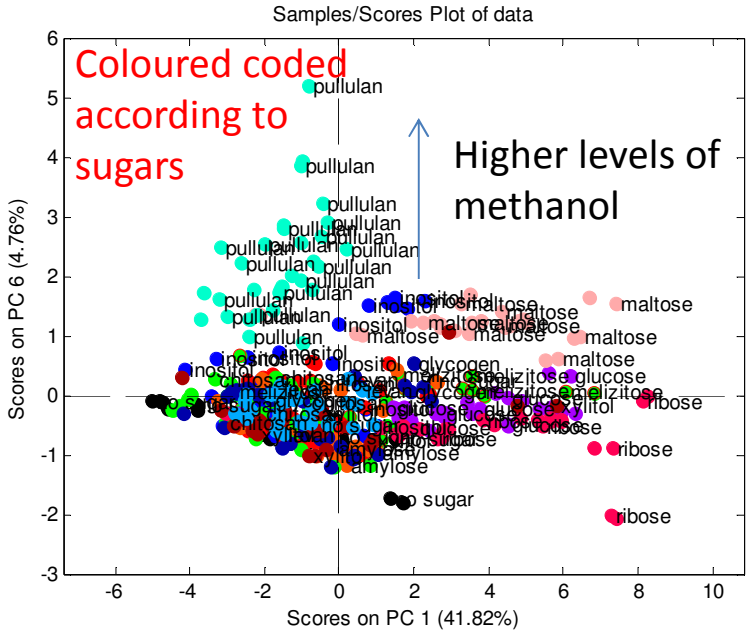
PC2 reflects an increase in trimethylamine and dihydrouracil levels for clade 3

Multivariate analysis on ^{1}H NMR data

PCA on 428 ^{1}H NMR spectra

PC3 (not shown): separation due to adenine levels





PC4: separation due to propionate levels

PC6: separation due to methanol levels

PC5 (not shown): separation due to maltose levels

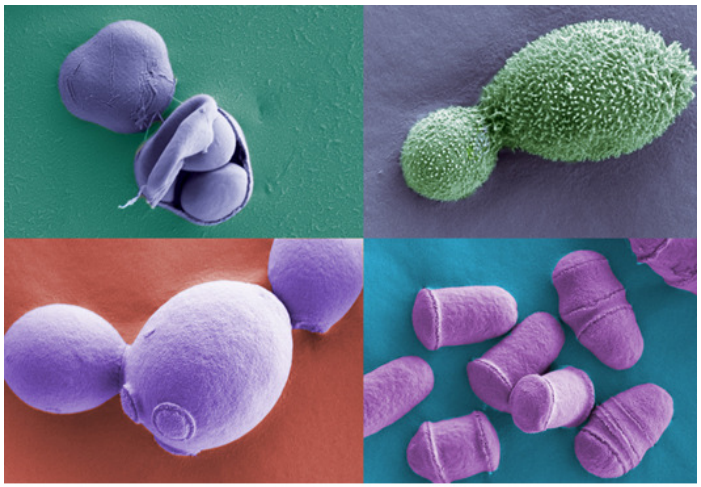
PC7 (not shown): separation due to propanediol levels

Acknowledgements

Mike Peck, Diana BenaventCortes, Sandra Stringer, Martin Webb, Institute of Food Research

The National Collection of Yeast Cultures (NCYC) at IFR

The National Collection of Yeast Cultures (NCYC) is one of the largest yeast collections in the world, making it a valuable resource for academics as well as industry

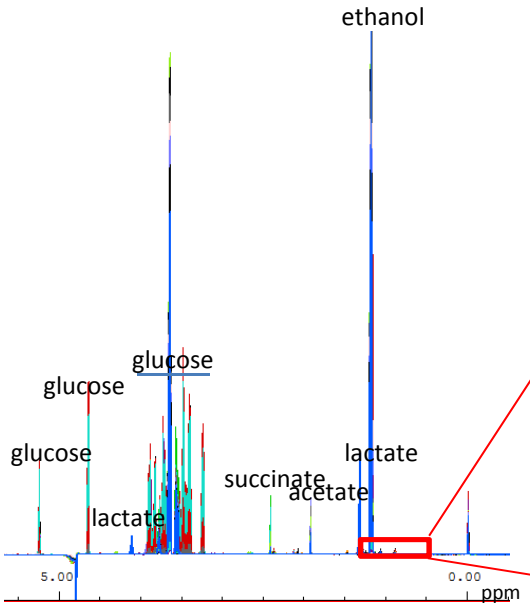


yeast diversity

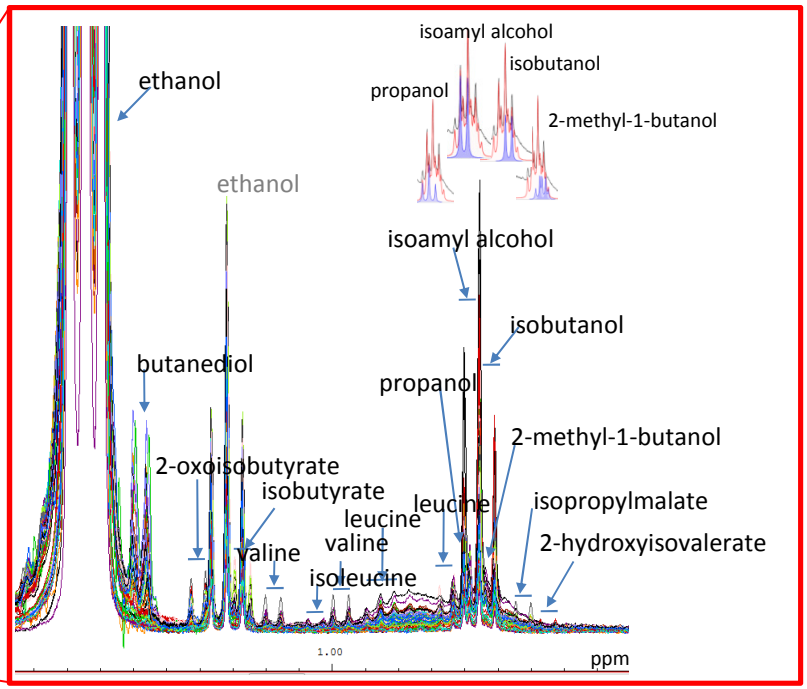
- **One of the research projects at IFR is sequencing the genomes** of a collection of yeast strains, to help unlock the great biodiversity within yeasts to produce biofuels and other chemicals more sustainably
  - We also decided to **screen metabolites** produced by the yeast strains

^{1}H NMR spectra of spent medium

Overview



Region 0.70-1.70 ppm



- The spectra are dominated by signals from ethanol, glucose (substrate), lactate, succinate and acetate
- But many low levels metabolites are also detected
- In total 61 metabolites were found in yeast spent medium
- The entire NCYC collection was screened by ^{1}H NMR (3552 spectra)

Aim1: detecting high value bio-based chemicals

National Renewable Energy Laboratory

Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

Top Value Added Chemicals From Biomass

Volume I: Results of Screening for Potential Candidates from Sugars and Synthesis Gas

August 2004

This report, the first of several envisioned to examine value-added products from all biomass components, identifies a group of promising sugar-derived chemicals and materials that could serve as an economic driver for a biorefinery. By integrating the production of higher

Building Blocks
1,4 succinic, fumaric and malic acids
2,5 furan dicarboxylic acid
3 hydroxy propionic acid
aspartic acid
glucaric acid
glutamic acid
itaconic acid
levulinic acid
3-hydroxybutyrolactone
glycerol
sorbitol
xylitol/arabinitol

in red: seen in the yeast spent medium

30 Hottest Molecules of 2016

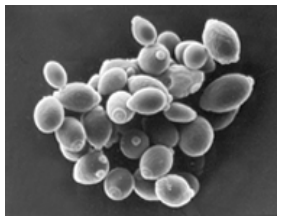
From acetic acid to xylose. What's Hot, What's Not.

Rank	Molecule	Some appeal	Real value now	Major opportunity
1	Dextose/fructose/glucose (C6 sugars)	93%	62%	31%
2	Polyhydroxyalkanoates (PHA)	74%	52%	30%
3	Lactic acid	93%	72%	28%
4	Omega-3 DHA	88%	64%	25%
5	Ethane/ethene	68%	39%	25%
6	Hexamethylene diamine (HMD)	68%	57%	25%
7	Isoprene	79%	62%	24%
8	Ethanol	79%	53%	24%
9	Polyethylene terephthalate (PET)	83%	57%	23%
10	Omega-3 EPA	86%	62%	23%
11	Limonene	81%	38%	23%
12	Hydrogen	84%	58%	22%
13	Ethylene	64%	43%	21%
14	Renewable jet fuel	90%	55%	20%
15	Succinic acid	87%	64%	20%
16	Polylactic acid (PLA)	85%	63%	19%
17	Polyethylene furanoate (PEF)	79%	49%	18%
18	Xylose/arabinose/galactose (C5 sugars)	92%	59%	17%
19	N-butanol	89%	49%	17%
20	Renewable diesel (e.g. 9-15 carbon organics)	90%	54%	17%
21	Renewable natural gas (e.g. methane, CNG)	78%	40%	16%
22	Isobutanol	90%	56%	15%
23	2,5 furan dicarboxylic acid	70%	34%	15%
24	Carbon monoxide	81%	40%	15%
25	Estolide	29%	14%	14%
26	Propane/propene	61%	39%	14%
27	Butanediol (BDO)	90%	61%	14%
28	Biodiesel (i.e fatty acid methyl esters)	83%	50%	14%
29	Citric acid	83%	52%	14%
30	Hexanediol	83%	48%	14%

http://www.biofuelsdigest.com/bdigest/2016/01/13/the-30-hottest-biobased-molecules-for-2016/31/

Aim 2: Producing biofuel and added value products from biomass

Biorefinary at IFR:
Fundamental science and knowledge to process biomass conversion (including food waste) to produce fuels, power, heat, and value-added chemicals from biomass



- **Biofuels**
- **high value products**

Acknowledgements

Adam Elliston, Keith Waldron, Keith Roberts, Institute of Food Research

Last words

Metabolomics is

- **Versatile**: any type of sample, only 400 μl or 20 mg of material needed
- **Quick**: NMR, a profile in 10 min for the main (60-100) compounds; MS, a profile in 30 min for 20-400 compounds
- NMR, **great for screening large cohorts of samples**, high through-put automation; MS, very sensitive, and fairly automated
- The metabolite identification is expertise-based but once the list of IDs is there, some degree of routine analysis can be implemented
- **Quantitative**: easier to compare with the literature

We thank BBSRC for funding
BBSRC bioscience for the future