Metabolite Analysis and Mass Spectrometry

Metabolites Are the Canaries of the Genome



A single base change can lead to a 10,000X change in metabolite levels

Metabolism is "Understood"



Metabolite

- Metabolites are the intermediates and products of metabolism
- Primary metabolites are synthesized by the cell because they are indispensable for their growth
- Secondary metabolites are compounds produced by an organism that are not required for primary metabolic processes
- A by-product of the breakdown of either food or medication by the body
- Any compound detected in the body <1500 Da

Metabolites

Primary/central metabolites

Nucleosides Nucleotides Purines Pyrimidines Coenzyme As Pyridoxals Biotins Quinones Nucleoside sugars Flavins Pteridines Tetrapyrrols Thiazoles Amino acids Acyl amino acids Amino acid phosphates Aromatic acids Alcohols and polyols

Monosaccharides Aminoglycosides Sugar phosphates Alcohol phosphates Acyl phosphates Sugar acids Sugar lactones Deoxysugars Phospholipids Glycerolipids Sphingolipids Glycolipids Carotenoids Fatty acids Hydroxy acids Keto acids Dicarboxylic acids Tricarboxylic acids

Secondary/remote metabolites

Indoles Alkaloids Aminoketones Cyclic amines Aromatic amines Cobalamines Polyamines Polyenes Polyethers Polyphenols Dideoxysugars Sugar sulfates Disaccharides Trisaccharaides Tetrasaccharides Glycopeptides Glycosylglycerides Glycosylated natural products Prostanoids Steroids Sulfolipids Leukotrienes Lipoamides Polyketides Terpenes Macrolides Bile acids

The Metabolome is Connected to All Other "Omes"

- Small molecules (i.e. AMP, CMP, GMP, TMP) are the primary constituents of the genome & transcriptome
- Small molecules (i.e. the 20 amino acids) are the primary constituents of the proteome
- Small molecules (i.e. lipids) give cells their shape, form, integrity and structure
- Small molecules (sugars, lipids, AAs, ATP) are the source of all cellular energy
- Small molecules serve as cofactors and signaling molecules for both the proteome and the genome
- The genome & proteome largely evolved to catalyze the chemistry of small molecules

Metabolomics to Phenotype



Metabolite Analysis

- Metabolomics: quantitative measurement of all low molecular weight metabolites in a given sample, cell or tissue.
 - Quantitation of specific metabolite.
 - Profiling a group of related compounds.
 - Quant/qual of all metabolite.
 - fingerprinting: rapid global analysis of a class of compounds.





01196 6426/00

01150 3/3/0





01196 6426/02

Challenges in Metabolomics

=

•Metabolites have a wide range of molecular weights and *large variations* in concentration

•The metabolome is much <u>more dynamic</u> than proteome and genome, which makes the metabolome more *time sensitive*

•Metabolites can be either polar or nonpolar, as well as organic or inorganic molecules. This makes the *chemical separation* a key step in metabolomics

•Metabolites have <u>chemical structures</u>, which makes the **identification** using MS an extreme challenge



Metabolomics Flow Chart



R. Gooadacre et al Analytica Chimica Acta 879 (2015) 10-23



Quantitative (Targeted)



Metabolite Identification & Quantification

Profiling (Untargeted)



Sample Prep



Metabolite Identification

a Targeted metabolomics



Nature Reviews | Molecular Cell Biology

Nature Reviews Molecular Cell Biology 13, 263-269 (April 2012)



Figure 1. General workflows of metabolomics study, including study design, sample preparation, data acquirement and data mining.

Published in: Zhi-Gang Gong; Jing Hu; Xi Wu; Yong-Jiang Xu; *Critical Reviews in Analytical Chemistry* Ahead of Print DOI: 10.1080/10408347.2017.1289836 Copyright © 2017 Taylor & Francis Group, LLC

Study Design

Hierarchical Pilot Study to determine variation



Trutschel, D., Schmidt, S., Grosse, I. et al. Metabolomics (2015) 11: 851.

Knowledge discovery in metabolomics: An overview of MS data handling



Journal of Separation Science <u>Volume 33, Issue 3, pages 290-304, 19 JAN 2010 DOI: 10.1002/jssc.200900609</u> http://onlinelibrary.wiley.com/doi/10.1002/jssc.200900609/full#fig1

Study Design



Trutschel, D., Schmidt, S., Grosse, I. et al. Metabolo mics (2015) 11: 851.

cost, sample availability and the accuracy of the tests

Study Design

- Internal Standards?
 - Recovery standards to estimate the % recovery from sample preparation method
 - Analytical standards for normalization in data analysis. Must be immune from ionization/detection variation and not affect sample features. Easier for metabolite subsets than for untargted.



Figure 1. General workflows of metabolomics study, including study design, sample preparation, data acquirement and data mining.

Published in: Zhi-Gang Gong; Jing Hu; Xi Wu; Yong-Jiang Xu; *Critical Reviews in Analytical Chemistry* Ahead of Print DOI: 10.1080/10408347.2017.1289836 Copyright © 2017 Taylor & Francis Group, LLC

Sample Preparation Method Criteria



Vuckovic, D. Anal Bioanal Chem (2012) 403: 1523.

Sample Preparation Method Comparison



78:743

	Metabolo me coverage	Free concentratio n	Metabolism quenching step	Repeatabilit y	In vivo sampling	Ionization suppression	Protein removal efficiency	Small volume samples
Solvent precipitation extraction	or Excellent but depends on solvent	No	Limited and only if rapid after collection	Excellent	No	High	Low	No
Microdialysi s	No data	Yes	Yes	No data	Yes	High	High	Yes—no sample is withdrawn
Ultrafiltratio n	Excellent for polar metabolites; poor for hydrophobic metabolites	Yes	Limited and only if rapid after collection	Good	No	High	High	No
							-	
SPE	Good	Depends on sample pre- treatment	No	Good	No	Medium	Not evaluated to date	No
	I		1	1	1		1	
DBS	Good	No	No	Medium	No	Not evaluated to date	Low	Yes
TFC	Good	Not known	No	Medium	No	Medium	Not evaluated to date	Yes
SPME	Good but lower sensitivity	Yes	Yes (in vivo)	Good	Yes	Low	High	Yes—no sample is withdrawn for in vivo SPME

Vuckovic, D. Anal Bioanal Chem (2012) 403: 1523.

Sample Normalization



Sample Normalization

- Pre-acquisition
 - Adjust volume or weight (can input for data processing.
 - Measure concentration of known unaffected metabolite (creatine in urine).
 - Normalize to some measure osmolality, UV...





Metabolomics Methods

Technique	Advantages	Disadvantages
GC–MS	 High resolving power and accuracy Reproducible retention time Applicable to volatile and semi-volatile compounds (through derivatization) low cost compared to LC-MS and NMR Comprehensive databases available 	 Volatility can be a restriction Heat sensitive compounds cannot generally be analysed Derivatization may complicate sample preparation and identification (due to additives and multiple derivative products)
LC-MS	•High accuracy <comma> resolving power<comma> sensitivity and specificity•Sample preparation is minimal compared to GC–MS (generally no derivatization required)•Applicable to complex mixtures<comma> polar and non- polar compounds</comma></comma></comma>	•Without comprehensive MS–MS or MSnstructural information is limited•Matrix effects•Formation of multiple adducts
NMR	 Robust and highly reproducible Provides very specific structural information Non- destructive (samples can be recovered) Minimal sample preparation Highly quantitative 	•Very expensive•Spectral interpretation is time consuming•Low sensitivity (micromolar range) compared to MS (picomolar range)
FT-IR	 Rapid and high throughput Relatively inexpensive Provides information rich data 	•Water is an issue in mid IR (samples must be dried) <comma> although wet samples can be analysed using attenuated total reflectance (ATR)•Mixtures may complicate data interpretation•Not all compounds can be detected•Low chemical specificity 31</comma>

R. Gooadacre et al Analytica Chimica Acta 879 (2015) 10-23

NMR or MS Profiling	Very Robust	High Throughput	500- 10,000 variable s	5-50 identified compounds
GC-MS	Robust	High Throughput	500- 1000 primary metabo lites <550 Da	50-200 identified compounds
LC-MS	Not as Robust	Low Throughput	50-70 second ary metabo lites <2500 Da	20-100 identified compounds

Metabolomics Separation methods

- Gas chromatography (GC)
 - one of the most widely used and powerful methods
 - high chromatographic resolution
 - compounds must be volatile (or derivatized)
- High performance liquid chromatography (HPLC)

- lower chromatographic resolution

- wider range of analytes can be analyzed (polar)
- Capillary electrophoresis (CE)
 - higher theoretical separation efficiency than HPLC
 - suitable for wider range of metabolites than GC
 - most appropriate for charged analytes (electrophoretic technique)

Metabolite Analysis

GC/MS

Alkylsilyl derivatives Eicosanoids Essential oils Esters Perfumes Terpenes Waxes Volatiles Caratenoids Flavenoids Lipids

Alcohols Alkaloids Amino acids Catecholamines Fatty acids Phenolics Polar organics Prostaglandins Steroids

overlap

Less Polar

More Polar

LC/MS

Organic Acids

Nucleosides

Ionic Species

Nucleotides

Polyamines

Organic Amines

GC-MS Chromatogram of a Biological Mixture



HPLC of a Biological Mixture


NMR

- Advantages
 - No sample separation necessary
 - Essentially universal detector
 - Non-destructive
- Disadvantages
 - Low sensitivity
 - Results difficult to interpret
 - Decreased quantification
 - Expensive



Figure 1. General workflows of metabolomics study, including study design, sample preparation, data acquirement and data mining.

Knowledge discovery in metabolomics: An overview of MS data handling



Data Pre-Processing

- Noise reduction
- RT correction
- Feature finding
- Integration
- Chromatogram alignment (between all samples)

Data Pre-Processing



Enhancement of signal to noise ratio

- elimination of noise (wavelets)
- background removal (e.g. PALS)





Synchronization of time axes

HPLC chromatograms of green tea extracts







Data Pre-Treatment



Treatment Missing Values

- Discard all
- Replace by average value (A)
- Replace by average of nearest neighbors (B)



Normalization

- Scalar
 - Assumes total signal per sample is =
- Internal Standard
 - Ratio response metabolite/ISTD
 - Optimal combination of ISTDs, endogenous
- Pooled QC based
 - Batch/analyte ratio (mean, median)

Centering, scaling, transformation

Class	Method	Formula	Unit	Goal	Advantages	Disadvantages
I	Centering	$\tilde{X}_{ij} = X_{ij} - \tilde{X}_i$	0	Focus on the differences and not the similarities in the data	Remove the offset from the data	When data is heteroscedastic, the effect of this pretreatment method is not always sufficient
п	Autoscaling	$\tilde{X}_{ij} = \frac{X_{ij} - \overline{X}_i}{S_i}$	(-)	Compare metabolites based on correlations	All metabolites become equally important	Inflation of the measurement errors
	Range scaling	$\tilde{\mathbf{X}}_{ij} = \frac{\mathbf{X}_{ij} - \overline{\mathbf{X}}_i}{\left(\mathbf{X}_{i_{max}} - \mathbf{X}_{i_{min}}\right)}$	(-)	Compare metabolites relative to the biological response range	All metabolites become equally important. Scaling is related to biology	Inflation of the measurement errors and sensitive to outliers
	Pareto scaling	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}}$	0	Reduce the relative importance of large values, but keep data structure partially intact	Stays closer to the original measurement than autoscaling	Sensitive to large fold changes
	Vast scaling	$\tilde{X}_{ij} = \frac{\left(X_{ij} - \overline{X}_i\right)}{S_i} \cdot \frac{\overline{X}_i}{S_i}$	(-)	Focus on the metabolites that show small fluctuations	Aims for robustness, can use prior group knowledge	Not suited for large induced variation without group structure
	Level scaling	$\tilde{x}_{ij} = \frac{x_{ij} - \overline{x}_i}{\overline{x}_i}$	(-)	Focus on relative response	Suited for identification of e.g. biomarkers	Inflation of the measurement errors
111	Log transformation	$ \begin{split} \vec{x}_{ij} &= ^{10} \log \Bigl(\times_{ij} \Bigr) \\ \vec{x}_{ij} &= \vec{x}_{ij} - \vec{\bar{x}}_i \end{split} $	Log O	Correct for heteroscedasticity, pseudo scaling. Make multiplicative models additive	Reduce heteroscedasticity, multiplicative effects become additive	Difficulties with values with large relative standard deviation and zeros
	Power transformation	$\tilde{x}_{ij} = \sqrt{\left(x_{ij}\right)}$ $\bar{x}_{ij} = \tilde{x}_{ij} - \tilde{x}_{ij}$	√0	Correct for heteroscedasticity, pseudo scaling	Reduce heteroscedasticity, no problems with small values	Choice for square root is arbitrary $\mathop{\mathrm{g}}\nolimits_{\mathop{\mathrm{S}}\nolimits}$

Centering, Scaling and Transforming

- A. RawB. CenteringC. Auto
- D. Pareto
- E. Range
- F. Vast
- G. Level
- H. Log
- I. Power
- All Feature-wise normalization methods



Statistical Analysis

- Univariate
 - Preliminary overview of data
 - T-test, ANOVA
 - Group-wise differences parameter by parameter
 - Reduces to analytes which show the strongest responses (+/-) to tested conditions
- **Multivariate**
 - In addition to differences from conditions also looks at relationships among analytes



Comput Struct Biotechnol J. 2013; 4

Statistical Analysis

- Clustering
 - Group and visualize samples according to intrinsic similarities in their measurements, irrespective of sample groupings
 - Multivariate
 - Hierarchical (HCA) or partial clustering algorithms (k-means

Hierarchical Clustering



Heat map

Statistical Analysis

- Classification
 - Based on training set create model of data
 - Assigns a new sample to a category and its metabolites to known group or pathway
 - Need to validate model
 - ROC, Crossvalidation

Data Analysis Progression

- Unsupervised Methods
 - PCA or cluster to see if natural clusters form or if data separates well
 - Data is "unlabeled" (no prior knowledge)
- Supervised Methods/Machine Learning
 - Data is labeled (prior knowledge)
 - Used to see if data can be classified
 - Helps separate less obvious clusters or features
- Statistical Significance
 - Supervised methods always generate clusters -- this can be very misleading
 - Check if clusters are real by label permutation

Name	Main Application	Specific Features	License	User Interface
metaXCMS	Importing XCMS output	Post processing of XCMS for comparison of multiple (≥3) classes and visualizing statistical analyses.	Free	R language ^{≛)} and GTK
XCMS2	Importing tandem mass spectrometry (MS/MS) raw data	Processing of tandem mass spectrometry data for metabolite identification and structural characterization	Free	Plug-in of R language <u>*)</u>
MetAlign	Importing many common formats, including Masslynx, Xcalibur, netCDF, and the old-style HP/Agilent format of GC- MS / LC-MS data	Interface-driven data processing program. Includes baseline correction, smoothing, feature detection and alignment	Free	Local application (GUI)
MAVEN	Data processing of LC- MS and pathway visualization	Tools for all aspects of data analysis, from feature extraction to pathway-based graphical data display	Free	Local application (GUI)
metaP-server	Statistical analysis, database searching, pathway visualization	A web-based metabolomics data analysis tool	Free	Web
MetDAT	Statistical analysis, database searching, pathway visualization	A modular and workflow- based free online pipeline for mass spectrometry data processing, analysis and interpretation	Free	Web
Pathway projector	Pathway visualization	A Web-based zoomable pathway browser that uses KEGG atlas and Google Maps API	Free	Web
		A web-based		





Published in: Zhi-Gang Gong; Jing Hu; Xi Wu; Yong-Jiang Xu; *Critical Reviews in Analytical Chemistry* Ahead of Print DOI: 10.1080/10408347.2017.1289836 Copyright © 2017 Taylor & Francis Group, LLC

Data Interpretation

- Identification of Significant Metabolites
- 4 levels of ID
 - Positively identified compounds
 - Confirmed by match to known standard
 - Putatively identified compounds
 - Match to MS + RT or MS/MS + RT
 - Compounds putatively identified in a compound class
 - Unknowns

Databases for Metabolomics

- NMR spectral databases
 - Primarily small molecule spectra, not all metabolites
- MS or MS/MS spectral databases
 - Primarily small molecule spectra, not all metabolites
- Compound databases
 - Mostly compound names, structures, IDs, physprops
- Pathway databases
 - Mix of metabolite, drug, protein, signaling pathways
- Comprehensive metabolomic databases
 - Combines most/all of the above, focus on metabolites

Databases for Metabolomics

- NMR spectral databases
 - Primarily small molecule spectra, not all metabolites
- MS or MS/MS spectral databases
 - Primarily small molecule spectra, not all metabolites
- Compound databases
 - Mostly compound names, structures, IDs, physprops
- Pathway databases
 - Mix of metabolite, drug, protein, signaling pathways
- Comprehensive metabolomic databases
 - Combines most/all of the above, focus on metabolites

NMR Spectral DBs

SBDS

660 ×	11 RO DR Spectral Instalance for Organic Composition, 2000				
(#1.8.1.8.1.8.1.8.1.8.1.8.1.8.1.8.1.8.1.8	tees and go.g. tells. top-loc. Definition () op	k forget			
() The locate to divertising Assession of the	many defensions man banks. The base for Multimode, Author Balls, all log	or Alph longs and lutters B			
Organic Compounds \$085		CD .497			
Welcome to Sp The truthe strengers	ectral Database for Organic Compounds, SDB	S. (
MMR 7 Salts, 4 Hayannes, M	Transgouves and D. Tercenoto				
EDR & Burners					
Renar K Tanata and J Ministr	and C.Samure				
What's form		0			
1000-00-00 Review with the detect	2 (2004) VICE (R.M 2004) (2.08 (4.00))				
, 2008 (0.21), here all in deat	(J00,95,9,10.0) J00(01,0) J0001				
The second					
We and change our later the comparison We called the comparison to the later of the contract of the comparison processing of the comparison of the compariso	Topic surdry, landscore, Proposel, Price and no surf- interfaced and frain any prices of minimum Partic and share on the Lands for any landscore that some work from any or in the database. We inform an any fire form and only on the transmit frain the Constant Society intermined and any	5 I I I			
Access to the detailague is free of which or companying other same in profile challing a contribution of or	Harge, Report of much robot to be dealers for a dealers for the local set of the life general line life, if annual of the set of the life set of the set of the life set of the set of t	ų.			
We also request that after you at	respet consult of and maxima the interded usage is propine of the IOM. In the data of our IOM in your publication or preservation, is proper advectedaptment by p				
: Cont :	is not access an insertional lists by using automated processes. Next Longeural and Coston settings if you can not present to each page.				
	He had a summittee access from one publicate site				
	MANAOD				
	MINCD				
0.0.0	Madege Meddelarity Conservat Database Med 31				
ALL A R. M. A Composition	and and and the second se	to the second second second second			
A.A. NUMBER DO	and support the last of the support of the last of	a sea man an anna a			
	Madison Metabolomics Consortium Databe	64°			
INCOME Y ADDRESS AND Y DESCRIPTION	TRANSPORT AND TRANSPORT TRANSPORT	100.0			
The second secon					
()	P	(1044			
Text_based Search		+ BARICE - HORE			
Tapat process	in Metabolomics Consortium Database	+ MARCO TONS			
Easth: (http://di.doi.org/10.01/au.met //) +Strate MACD					
A suppose (bits to 'same' search)		- 000703			
NAME NAME THAT IS ADDRESS OF	Bosoniae, and J. J. Markey, "Balancin derification on the Markey	- SECOND			
(TRANSPORT)	the second second a second	1907 201 101 101 101 101			
	Introduction	(Data Burbelles			
- A DESCRIPTION OF A DE	is namened by the National Magnetic Resonance Facility at Hadson, is a	internation (1930)			
CONTRACTOR OF STREET, ST.	while the goal of the bindhair bindhairman Consertant Debutter and AMCO has	Comparents			
MADE Search Second The	troughput MMI and MI approaches to the identification and puerification of cliningical samples. MMICS currently serves as a had-to information on unrail	inenant and			
Name and Address of the Owner o	a Harwel gallered from machines and the scientific freedom. Each	agarinana			
and and a second	evital familie, rames and synamics, situation, physical and chemical	Constantine and American			
Google + (heat)	Internet for any compounds under definiel conditions where analysis, MAN small to empirical and/or Technitical approaches, calculated techsporter	with MACO			
	are the presence of the instabular in officient temporal spaces, and extension process and other public depleteres, such as the Turth Exception of Second	Panettai III			
	and "Address. The MMCD search angree augusts vename late mong and	NAME AND ADDRESS OF AD			
	Construction of the second state of the second of the second state	and the second se			

NMRShiftDB



BMRB



MS Spectral DBs

NIST/AMDIS



Metlin



MassBank



GolmDB

0.0		GMID - the Golyn Metabolome Database	
- 0 0	b 🕼 🕍 gend maima-goles mags	64	6 Literate
III Geogle	Apple iCloud Facebook Twitter	Witigedia Valuo News * Popular *	
herne		OLM METABOLOME DATA	
search ma analysis Golfiefspace profile data decision trees data entities web services download	The Goins Hetabolisme Databas of reference mass spectra from chromotography (GC) coupled the ammg at the identification of con- the comparison of observed mass such as the GHO. The GHO comp reference substances and freq spectrum linked to chromotograph	E10HD) facilitates the search for and dissummation biologically active metabolies quartified using gas mass spectrovery (1915). GC/UB profiling studies spectra and recently (1915). GC/UB profiling studies spectra and recently (1915). GC/UB profiling clusters inter mass spectra and recently interance bizanies rank spectra and recently interance bizanies control of gas and the spectra big profiling the activity observed mass spectral tags (1951) mass in control of gas underhilded metabolities.	Mass Spectrum of Alanine (27MS)
blog thanks to publications imprivit (P contact		Many current approaches in metabolomics are charact quantitative methods, similar to the previous develop application of state-of-the-art high-throughput teth complexity of the generated data leading to an increas annotation and data mining.	ented by an ongoing transition from qualitative towards ment in genomics, transcriptionics and proteonics. The nologies results in a significant growth in size and and demand for computational methods of visualisation.
help	miss experiment metabalite profiling such, the GMD constitutes a store methods based on the analytical r	The GMD incorporates quantitative data of metabolite ge system for data on the molecular level, providing acc esuits.	pool size changes. As ass for computational
	Arabidopsis	The GMD facilitates orthogonal metabolits profile multiple experimental conditions or to profile all m unique data set as a reference, the GMD classifier results and experimental factors.	ng; this is to say, to profile a single metabolite cross etabolites within a single experimental setup. Using this s unknown metabolic signatures with respect to species
	Your use of this site is governed b	y our <u>Terms of Use</u> linked below.	
	Punotinis		
	This project is supported and fur	ded by the <u>Has Planck Society</u> .	search Foundation DFG and the Europe

Compound DBs

ChEBI



PubChem



ChemSpider



Ligand Expo 4 > 4 0 00 A A C 3,http://ligand-expo.ttsb.org/id-set PDE Ligand Expo **Chemical Component Search Tools** Use the forms below to search for chemical components within the PDB Component Dictionary. Search for chemical components by 3-letter component identifier code, molecular name, molecular formula, SHILES description, or InChi/InChiKey chemical descript IncNVMCNKey chemical description. Toos can also check to see it a 3-letter code is being held by a deposition in progress. • Chemical substructure services can also be conducted by attering from a chemical drawing created within the Marvindleatch tool. • Either services and the description or chemical data fill (we draw more for executable formatic), or draw a 2D chemical substructure from sortable description or chemical data fill (we draw more for executable formatic), or draw a 2D chemical substructure from sortable description or thread data fill (we draw more for executable formatic), see a list of POD codes, or to Search for instances of a chemical component thready the POD. The **Biologica** polate dates you to simple we a list of POD codes, or to the history to be also also polar and mark() formatic. • Too can be instanced and the code is also be also polar drugs, and common aromatic ring systems by using the Browee feature in the too means. Browse feature in the top menu bar Your query results are also searchable! Each hit from your initial query will contain links to continue searching by similar name, chemical Interfactor PEARS IN HOLECULAR NAME, FORMULA, AND DESCRIPTOR SEARCH OPTIONS Search term Search type Component identifier (3-tetter code: 4) 6e SKETCH INPUT AND/OR STRUCTURE SEARCH OPTIONS File name (choose File) no the selected File format MOL/30F format 1 -- 08 ---Display size (Small 4) SMILES string

Other Compound DBs

3DMet

000	SDMET: A three-dimen	sional encours database of n	atural metabolite			
A + + 0 + A	A 6 the step //www.benet.doa.aftic.go.jp/			0 M.O.	Congra	_
(1) Genome Alber 1 Platform	Drug ADME in Outabors: PolySearch- rch Results	Ligand Expo Develoads Pro	es-Drugsoffution	"Prous Scien	Cox Bination are Limited	30
itana	A three	3DM e-dimensional- of natural m	ET structure retabolite	database s		
Search Molecules	<u>3DMET</u> is a database collecting three	e-dimensional structure	s of natural m	etabolites.		
Documents	Three-dimensional structure is an important factor for molecular recognition. This recognition is always performed in the living cell. As we know, asome shrinkly is also essential factor of the 5D-structure for molecular recognition. However, many programs of 2D-3D convertes can net transform correctly about all of the narraw metaolisms. Thus, 3DMET 1O surversare translations of manufacture interdibutes (has been developed.					
What's new? This database has not only 3D-structures but also some descriptors. Each structures are searchable by these descriptors. This system also has simple search system for substructures to the molecule. The search can be performed by three levels depending on the accuracy.					ors. This system also has a eccuracy.	
Contact us				1	Foday's molecule (B00179)	
Q&A	The benefits to use 3DMET are:					
	 we can view a minimized 3D-structure the 3D-structures are verified. we can refer both of the 2D and 3D st we can obtain many molecular descrip we can find any molecular selated to b -> see "Documents" 	e of the molecule. nactures of a molecule. tors for 2D and 3D structure roth of descriptors and subset	s. actures.		60000	
					Jimol	
	To showing 3D-structures, Imol is used in the molecule", you may need to install the Jana J	e each data page. If you cans Rantime Environment-	not see the structu	re of the 'Today's		

KNApSAcK

Link to Top pag http://consyn.ht	t at jokospisack "jopkop henti				
Incorporation to http://katorys.to Here, (http://an	program: 1. gebrapsack: japtete japtesamer(in 1. be selected from one of the following	tideword=(kayword) words, 'organism', 'nasiabolite',	Senate, 'C.D', and 'CALID		
4 Example I v Information on http://tamaya.te	he metabolite assigned to CODODOD1 (at printipaals, psycholopy fotoarao C	a C_20% care for retrieveal by D4 to only -C00000001			
e Enample 3 a The reported in http://kanays.te	uboline in Bacilias (an organises) can at gebragmack, jug/indo jug/inameeorg	te retrieved by attend word-bacility			
Words for organi strings.	ns or petabolism can be rekrieved by p	ero-iding at least three character	s hat forward matches with their		
CAUTION: (C) /	ry communicated in ENApEAcK do ing with KNApEAcK DB group (sta	infrase cannot be re-distributed or a insport(g), said, g().	and for assumential purposes by an	,	
Select by Organism (Metabolite () Molecular for	nula O C_ID O CAS_ID			
	fuelf	Clear			

MyCompoundID



LipidMaps

0.0	LIPID MARS LIPID Metabolites And Pathways Strategy			
+ 0 0 A A	C Ohtp://www.lipidmaps.org/	O A Q+ Google		
Genome Albe - 4 Platform Dru	g ADME n Database PolySearchrch Results. Ligand Expo Deventeads: Preus-Drugsoffsiture: "Preus Sciennce R	teraits GVK Execution and Limite		
00000	Sociated I Discussion I	Nexa I Eublications I Site M		
	About Lipid Classification Standards Experimental Data Databases Pathways	Tools Protocols Home		
Discussion Board	LIPID MAPS Highlights	Lipid Search		
ivided Tour of the LIRID MAPS Vidulta	Annual Meeting 2009 Lipidomics Impact on Cell Biology, Structure Montematy and Immunopathology May 6-7, 2009 - La Jolin, 2009	(Steer Common or Systematic Names, and Street, and (Steering) Advanced, Search		
Eath Anna	Guided Tour of the LPSD MAPS Website	Newsi and		
Schoolessenen Guessenensenenen Saterinaustanda Fitterin Landa Fitterin Landa Saterinaustanda Saterinaustanda Classification Lodates	Nethods is Enzymology Chapters Hethods in Enzymology, Academic Press, Val 432 New Devreal of USE 3925 Sectors in Enzymology 2015 Dates as a single FDF. Research A., Burysmith, H. W., Beerry Gertyn, R. C., Networks, R. and Devrealings of Entermines y Enterthies Meas	Hersbeite Labdemeins Shalles in Missenachapes, IAW 246, 2 refs brandel with Compacting, is Kis- land J. Brownie averages of biological and replaces and biological replaces and biological replaces and biological replaces and biological replaces and biological replaces and biological stategory, or search by ANOVA prostate. 1911078 - Subcellular		
Upidomics Meetings w 1810, 2001 - Landanian Markhala, 41 Supermental belogy 2000. Mark 1 Root - <u>LIPIC MAPS</u> Mark 1 Meeting, 2001 Landonwiss meast on Call Meeting, 2001 Directure Bechemistry and immunopathology.	Repetermently, in: Mol.: Express, R. Jackiewan Press, C.20, 19-12. Hoyk, T.D., Clark, B., Bromen, R. Jack, A. Wans, J. Lü, Natesa, Y. Zhao, and S. Sukramanian. Bailednermatike for Lipideneous (2007). Hoto: Express, Hoto, K. Jackiewan, H. D. Sukramanian, Bailednermatike for Convert, F. K., Quark, J. L., C. Lipideneous, C. S. D. Natesiries of Antibiotic Sciences, Advisories Press, S. J. Sokramanian, M. S. Sukramanian, M. S. Sukramanian, M. S. Sukramanian, M. S. Sukramanian, M. Sakramanian, S. Sukramanian, S. Sukramanian	Technomese, Madela, AM, AF, Calan and Calandaria and Calandaria replicates by analytic cohegory or replicates by analytic cohegory or search by Kei-2 Lool Algentral ratio in a given authoritikan bracks. EDISOLE - We have expanded four classification explore to include lipids from a variety of inclume lipids from a variety of inclume lipids from a variety of or classification courses tool or		
JPID MAPS is funded by an satisfield Institute of General	NeDenald, J. G., Thempson, B., NcDrum, E. C., and Russell, D. W. (2007) Estraction and analysis of starvals in Netherland matrices by Math-merformance limits' determination whetremous instruction more commenced.	MAPS structure detablace new contains a total of 10,789		

Other DBs

- 3DMet
 - 3D structure database of 8581 natural metabolites
- KNApSAcK
 - Database of 50,000 plant metabolites linked to species information
- MyCompoundID
 - Database of 11 million metabolically transformed metabolites
- LipidMaps
 - Database with 30,000 lipids (Fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterols, prenols, saccharolipids, polyketides)

Pathway DBs

KEG		ome.jp/kegg/)
A and a second sec	Course presence jointige Course presence jointige Course presence jointige Course presence pre	eter * (* * * * * * * * * * * * * * * * * *
	BioCyc/M	etaCyc
BIOCY	e Chud Facthook Tailter Wikipedia Yahoo News."	Needar *
BioCyc I BioCyc is a colle with tools for un Getting Starb New to BioCyc?	Database Collection ction of 3530 Pathway/Genome Databases (PGDBs), derstanding their data. ed Motical usage in:	
Select one click "chan every page substr. MG Search for the Search New User Ge	or more distabases (genomes) to search. To do so, ge organism database' in the box in the top right of . By default, BioCyc searches Excherichia coll K-12 1953. a gene or pathway using the Quick Search, or see menu for more options. adde >>	The same and the same and the same same same same same same same sam
		paused 1 2 3 4 5 6 7 8 9 10 11

BioCyc Databases

SMPDB (www.smpdb.ca)



Tweet to @mactorre

European Watercolar Biology Later

Tools

Pathway DBs

- Rich source of biological data that relates metabolites to genes, proteins, diseases, signaling events and processes
- Provide various tools to permit visualization and gene/metabolite mapping
- Often cover multiple species

Comprehensive MetDBs

KEGG





MetaboLights



HumanCyc



Comprehensive MetDBs

- Must contain >1000 metabolites
- Usually are organism specific
- Continuously updated
- Must contain
 - chemical + pathway data
 - or chemical + spectral + biological data
 - or chemical + pathway + spectral data
 - or chemical + pathway + spectral + biological data

Metabolite ID by GC-MS



Metabolite ID by GC-MS

- GC-MS is often best for identification of amino acids, organic acids, sugars, fatty acids and molecules with MW<500
- GC has higher resolution and better reproducibility than LC
- EI-MS is more standardized than soft ionization methods, so EI spectra are more comparable
- Most common route is to use AMDIS + NIST database

NIST 11 MS Database

- 243,893 EI spectra of 212,961 cmpds
- 9934 ion trap MS for 4649 cmpds
- 91,557 Qtof & QqQ spectra for 3774 compounds
- 224,038 RI values for 21,847 cmpds



NIST MS Search Software


Metabolite ID by LC-MS



Metabolite ID by LC-MS

- LC-MS is often best for identification of lipids, bases, amino acids, organic acids, fatty acids and other somewhat hydrophobic molecules
- Metabolite ID typically requires both MS and MS/MS data (along with retention time information) and internal standards
- Compound ID can be done by high accuracy mass matching and/or by MS/MS matching to spectral databases

Simple MW Search DBs

ChEBI (www.ebi.ac.uk/chebi/)



ChemSpider (www.chemspider.com)



PubChem (http://pubchem.ncbi.nlm.nih.gov/)

	The PubChem Project		
▲ ► (🖄 🔕 🛨 😳 pubchem.ncbi.nlm.nih.gov		C Reader
1 0	loogle Apple iCloud Facebook Twitter Wikipedia Yahoo News⊤ Popular⊤		I
	Databases > Upload Services > Help more >	f e t-s	
	Pub©hem	BioActivity Summary	8
	•	BioActivity Datatable	III
	BioAssay (2) 🚱 Compound (2) 🕕 Substance (2)	BioActivity SAR	88
	(I Go Advanced Search	BioActivity DataDicer	D
		Structure Search	2
	Try the new PubChem Search	3D Conformer Tools	ъ в о
		Structure Clustering	U
	Sole PubChemRDF is released. II allows PubChem data to be used with semantic web technologies and encodes PubChem information using the Resource Description	Classification	2
	Framework (RDF). Read more	Upload	8
	more 🔊	Download	ڪ
	Write to Heipidesk Disclaimer Privacy Statement Accessibility Data Citation Guidelines National Center for Blotechnology Information	PubChem FTP	(P)
	nor juli j		

HMDB (www.hmdb.ca)



Advanced MS Search DBs

- These databases support not only MW or MW range searches, but also support parent ion searches (positive, negative, neutral), peak list searches (from MS or MS/MS data) as well as MS/MS spectral matching
- These DBs are intended more for MSbased metabolomics and compound ID than the simple MW search tools

Advanced MS Search DBs

NIST/AMDIS (http://chemdata.nist.gov)



CFM-ID (http://cfmid.wishartlab.com)

800	CFM-ID	1
🔺 🕨 🚳 🔘 🕂 🔮 dmi	d.wishartiab.com	C Reader
C III Google Apple (Cloud	Facebook Twitter Wikipedia Yahoo News * Popular *	
CFM-ID ☆ Ut		
CFM-ID Competitive Fragme	entation Modeling for Metabolite Identification	
Welcome to CFM-I	D!	
CFM-ID provides a method i electrospray tandem mass s to produce a probabilistic ge techniques to adapt the moo Spectra Prediction:	for socurately and efficiently identifying metabolites in spectra generated by pectrometry (ISS-N453/K45). The program uses Competitive Pragmentation Modeling norative model for the MSMIS fragmentation process and machine learning ale parameters from date. This generated model can be used for: Preciding the spectra for a given chemical structure . This task predicts low/10/r, modum/02/r, and high=400 wereay MSMIS spectra for a minor structure provided model and used to the high=400 wereay MSMIS spectra for a minor structure provided model and the spectra of the high=400 wereage MSMIS spectra for a minor structure provided model and the spectra of the high=400 wereage MSMIS spectra for a minor structure provided model and the spectra of the spectra o	
Peak Assignment:	Annotating the peaks in set of spectra given a known chemical structure. This task takes a set of up to three input spectra (low/10/, madum/20/, and high/40/ energy levels) in peak list format and a chemical structure in SMLES or incN1 format, then assigns a putative imgenet annotation to the peaks in each spectrum.	
Compound Identification:	Predicted ranking of possible candidate structures for a target spectrum. This task tikes as of to to three proxy spectra (law)(VM, endury(VX), and (ng)(VM) energy levels) in possis list format, and ranks a lat of candidate structures according to how well they match the lipst spectra. This lit range be provided by the tusk or can be generated from HMDB or KEGG. The match is determined by matching the spectra for each candidate compound and computing a socie (laccard or DoPhotoLQ) based on the match between the predicted spectra and the input sociedra.	Hialina.

Metlin (http://metlin.scripps.edu/)



)- C) 🗙 🏚 🖪	http://www.massb	ank.jp/?lang	g⇒en			☆ ▼	Google	
/isited = Getting Started Latest Hea assBank High Resolution Mass	tdlines 55							
MassBank	ç			Н	igh Reso	ution Mas	Spectral D	vatabas
Database Service	S News	Manual	ana was unda	ted rame				
Statistics	May 12, 2011 Mar 25, 2011	Manual p Data (44	age was upda age was upda spectra) from	ited. rieu IMM, CAMS	& PUMC, C	nina was added		
Documents	Apl 05, 2011 Mar 24, 2011	Data (141 Documen	spectra) from t page was up	n PFOS rese odated. new	arch group v	ras added. rieu		
Download								All nev
Manuals	Data	base Se	rvice					
About MassBank	→ Spectrun Search	n	Quick Search	→ Peal Sear	ç ch	→ Substruct Search	ure → Adva Searc	nced :h
Contact	→ Spectral E	Browser	→ Batch S	Service	→ Brow	se Page	→ Record In	idex
Consortium Members	MassBank is fin	ancially sup	orted from N	ational Biosc	ience Datab	ase Center, Jap	an Science and	Technolo
	The Mass Spect	torometry S	Society of Jap	an officially s	upports Ma	isBank.		

MassBank (www.massbank.ip)

Metabolite ID by MS alone -Complications

- LC-ESI-MS often leads to the production of salt adducts, neutral loss species and multiply charged species
- Up to 80% of LC-MS signals arise from these "noise" sources
- Key challenge is to distinguish adducts or multiply charged species from parent ions or to group adducts or multiply charged species with parent ions

Adduct Formation



Effect on ESI Mass Spectrum

Sample Na Adducts

Neutral Loss Fragments

benzoic acid Nominal Mass = 122 Da



Handling MS Complications

- MZedDB, Metlin and HMDB are able to handle or predict adducts
- Metlin and MZedDB are able to handle or predict ion pairs or multiply charged species
- Metlin can potentially handle or predict neutral loss species
- Searching by MS or MS ranges can lead to lots of hits (high FP rate)

Resolving LC-MS Complications

- Identify, remove (or consolidate) adducts and multiply charged species
- Identify, remove (or consolidate) fragments (neutral losses, breakdown products, rearrangements)
- Identify, remove (or consolidate) isotope peaks
- Remove noise peaks (from sample blanks or peaks that do not appear in >2/3 technical replicates or peaks that do not show dilution trends in 4 dilution replicates)

Resolving LC-MS Complications

- Raw +ve mode spectrum
- Remove adducts
- Remove multiple charges
- Remove neutral losses
- Remove isotope peaks
- Remove noise peaks
- Final spectrum
- Repeat for –ve mode

- 15,000 features
- 12,000 features
- 10,000 features
- 8,000 features
- 3,000 features
- 2,500 features
- 2,500 M+H peaks
- 1,500 M-H peaks

Mzmine Met

MetFusion

MagMA

Exploiting High Mass Accuracy to ID Compounds

Туре	Mass Accuracy
FT-ICR-MS	0.1 - 1 ppm
Orbitrap	0.5 - 1 ppm
Magnetic Sector	1 - 2 ppm
TOF-MS	3 - 5 ppm
Q-TOF	3 - 5 ppm
Triple Quad	3 - 5 ppm
Linear IonTrap	50-200 ppm
	(10 ppm in Ultra-Zoom)

$$ppm = (\frac{m_{exp} - m_{calc}}{m_{exp}}) * 1E + 6$$

Formula Filters

 Use additional MS information (isotopic abundance) as well as chemical bonding restrictions (Lewis & Senior rules), known or presumed atomic compositional data and matches to known or hypothesized structures to reduce the possible # of structures/formulas that are generated

Impact of Mass Accuracy on Formula Numbers

Table 3

Number of possible molecular formulas at different levels of mass accuracy and the impact of isotopic abundance accuracy. A mass spectrometer capable of 3 ppm but with 2% correct isotopic pattern outperforms even a (non-existing) mass spectrometer with 0.1 ppm mass accuracy! The results are computed for randomly selected targets, so single results vary but the trend remains. LEWIS and SENIOR check was applied. Candidates with unrelated high element counts were already excluded

	with	out isotop	e abundan	ice inform	ation	2% isotopic abundance accuracy	5% isotopic abundance accuracy	
molecular mass [Da]	10 ppm	5 ppm	3 ppm	1 ppm	0.1 ppm	3 ppm	5 ppm	
150	2	1	1	1	1	1	1	
200	3	2	2	1	1	1	1	
300	24	11	7	2	1	1	6	
400	78	37	23	7	1	2	13	
500	266	115	64	21	2	3	33	
600	505	257	155	50	5	4	36	
700	1046	538	321	108	10	10	97	
800	1964	973	599	200	20	13	111	
900	3447	1712	1045	345	32	18	196	

Kind and Fiehn BMC Bioinformatics 2006 7:234 doi:10.1186/1471-2105-7-234

Some Points of Caution

- Many databases (PubChem, ChEBI, Metlin, FiehnLib, NIST) mix non-metabolites with metabolites or plant metabolites with animal and/or microbial metabolites or drugs/buffer reagents with metabolites
- This leads to many "silly" hits
- If you know the source organism use this information to limit the search or use organismspecific metabolome databases (HMDB, FooDB, DrugBank, KnapSack, etc.)

Pathway Analysis

Putting metabolomics data into a biological context



Biological Databases for Pathway Analysis

Database	Description	Website	Reference
Kyoto Encyclopedia of Genes and Genomes (KEGG)	466 pathways, 17,333 metabolites, and 9,764 biochemical reactions	<u>http://www.genome.jp/k</u> <u>egg/</u>	Kanehisa et al. (<u>2012</u>)
MetaCyc	2260 pathways from 2600 different organisms	http://metacyc.org/	Caspi et al. (<u>2008</u>)
The small molecule pathway database (SMPDB)	1,594 metabolites mapping 727 small molecule pathways found in humans	http://www.smpdb.ca/	Jewison et al. (<u>2014</u>)
WikiPathways	1,910 pathways	http://wikipathways.org/	Kelder et al. (<u>2012</u>)
Plant metabolic network (PMN/PlantCyc)	Multi-species pathway database for plant metabolomics	http://www.plantcyc.org/	Chae et al. (<u>2014</u>)

Metabolic Network Inference

*Search for the link between metabolome data and underlying metabolic networks.



* As an example: can we distinguish healthy from diseased networks:



From Spectra to Lists





Compand	Retention Time (min)	Canc.in Uine	Compared	Retention Time (min)	Canc.in Urine
Drie o ritosrito I como	0.02	401.*	Doc-Jo	635	- 4 m) - 25
Dris o mosmo I fermino	0.52	- 40.L	Dis-1-minorationin and	6.44	0.5
Dre advocine monorhoentrate	0.90	401	Das riporalis seid	650	0.5
Dis-anteie monprispize	100	401.		654	54
Dis-opiospioenaroannie Dis-ducosamine	106	22	Dris-cesationine	654	03
Drs_o_nhosnho_l_threanine	1.00	ഫ	Dos Jeu Pro	660	0.5
Dris-6-dimet hytamine runine	120		Dris-5-hwimowiksine	665	16
Dns-3-mellwl -hisirline	122	-01.	Dris-Cysine	673	160
Dos-tamine	125	834	Dns-N-nodeucine	681	0.1
Dns-carnosine	134	28	Drs-5.hvdinxydonamine	717	
Dns-Am	153	36	Dos-dimetry action	733	293
Dns-Asn	155	133	Dns-5-HIAA	746	18
Dos-twoctaurine	158	10	Dos-unbeliferone	7.47	19
Dos-homocamosine	1.61	39	Drs=23-daminoringingi acid	763	 - กา
Dos_cuaridoe	162	สาเ	Dns-L-omithine	770	15
Dns-Gin	172	633	Dos-4-aceiramidophenol	773	51
Dns-allantoin	183	38	Dns-grocaine	773	89
Dns-L-citruline	187	29	Dns-homocratine	776	33
Dns-1 for 3 -)-melhylihistamine	194	19	Dns-acelaminophen	797	82
Dns-adenosine	206	26	Dns-Phe-Phe	803	0.4
Dns-methylquanidine	220	-0L	Dns-5-melhyo xysalicylic acid	8.04	21
Dns-Ser	224	511	Dns-Lvs	8.16	184
Dns-aspartic acid amide	244	26	Dns-anine	8.17	⊲0L
Dns-4-hydroxy -proline	256	23	Dns-leu-Phe	822	0.3
Dns-Glu	257	21	Dns-His	8.35	1550
Dns-Asp	260	90	Dns-4-thialysine	8.37	⊲OL
Dns-Thr	3.03	157	Dns-benzylamine	8.38	⊲DL
Dns-epinephrine	3.05	-OL.	Dns-1-ephedrine	8.50	0.6
Dns-elhanolamine	3.11	471	Dns-tryptamine	8.63	0.4
Dns-aminoadipic acid	3.17	70	Dns-pyrydoxamine	8.94	⊲DL
Dns-Gly	3.43	2510	Dns-2-methyl-benzylamine	924	⊲DL
Dns-Ala	3.88	593	Dns-5-hydroxyliptophan	925	0.12
Dns-aminolevulinic acid	3.97	30	Dns-1,3 -diaminopropane	9.44	0.23
Dns-r-amino-butyric acid	3.98	4.6	Dns-putrescine	9.60	0.5
Dns-p-amino-hippuric acid	3.98	29	Dns-1,2 -diaminopropane	9.66	0.1
Dns-5-hydro xym elhyturicil	4.58	1.9	Dns-tyrosinamide	9.79	29
Dns-tryptophanamide	4.70	5.5	Dns-dopamine	10.08	140
Dns-isoguanine	4.75	-OL	Dns-cadaverine	10.08	0.08
Dns-5-aminopentanoic acid	4.79	1.6	Dns-histamine	10.19	0.4
Dns-sarcosine	4.81	7.2	Dns-3-melhoxy -tyramine	10.19	9.2
Dns-3-amino -isobutyrale	4.81	85	Dns-Tyr	10.28	321
Doc2-aminobudyrie seid	404	17	Dine eveloamino	10.44	- 401

From Lists to Pathways

	Retention	Canc.in		Retention	Canc. in
	Time	Unine		Time	Urine
Compound	(m n)	(M)	Campound	(m.m)	(MM)
Dris-o-phospho-L-seiine	0.92	<0.L *	Dns-le	6.35	25
Dns-o-phospho-L-tyrosine	0.95	⊲DL.	Dns-3-aminosalicylic acid	6.44	0.5
Dns-adnosine monophosphate	0.99	-40 L.	Dns-pipecolic acid	6.50	0.5
Dns-o-phosphoethanolamine	1.06	16	Dns-Leu	6.54	54
Dns-glucosamine	1.06	22	Dns-cystalhionine	6.54	0.3
Dns-o-phospho -L-threanine	1.09	⊲ 01.	Dns-Leu-Pro	6.60	0.4
Dns-6-dimet hylamine punine	1.20	⊲ 01.	Dns-5-hydroxylysine	6.65	1.6
Dns-3-melhyl -hislidine	122	80	Dns-Cysline	6.73	160
Dns-taurine	125	834	Dns-N-norleucine	6.81	0.1
Dns-carnosine	1.34	28	Dns-5-hydioxydopamine	7.17	⊲DL
Dns-Arg	1.53	36	Dns-dimelhylamine	7.33	293
Dns-Asn	1.55	133	Dns-5-HIAA	7.46	18
Dns-hypotaurine	1.58	10	Dns-umbelliferone	7.47	1.9
Dns-homocamosine	1.61	3.9	Dns-2,3-diaminoproprionic acid	7.63	⊲DL
Dns-guaridine	1.62	-dDL.	Dns-L-omithine	7.70	15
Dns-Gin	1.72	633	Dns-4-acelyamidophenol	7.73	51
Dns-allambin	1.83	3.8	Dns-procaine	7.73	8.9
Dns-L-cituline	1.87	29	Dns-homocystine	7.76	3.3
Dns-1 (or 3 -)-melhyihistamine	1.94	1.9	Dns-acetaminophen	7.97	82
Dns-adenosine	206	26	Dns-Phe-Phe	8.03	0.4
Dris-methylguaridine	220	.10⊳	Dns-5-methyo xysalicylic acid	8.04	21
Dns-Ser	224	511	Dns-Lys	8.16	184
Dns-aspariic acid amide	2.44	26	Dns-anine	8.17	⊲0L
Dns-4-hydroxy -proline	256	23	Dns-leu-Phe	822	0.3
Dns-Giu	257	21	Dns-His	8.35	1550
Dns-Asp	260	90	Dns-4-thialysine	8.37	⊲DL
Dns-Thr	3.03	157	Drs-benzylamine	8.38	⊲DL
Dris-epinephrine	3.05	.∎D	Dns-1-ephedrine	8.50	0.6
Dns-eihanolamine	3.11	471	Dns-tryptamine	8.63	0.4
Dns-aminoadipic acid	3.17	70	Drs-pyrydokamine	8.94	⊲DL
Dns-Gly	3.43	2510	Dns-2-methyl -benzylamine	924	⊲DL
Dns-Ala	3.88	593	Dns-5-hydroxyirpiophan	925	0.12
Dns-aminolevulinic acid	3.97	30	Dns-1,3 -diaminopropane	9.44	0.23
Dns-r-amino-butyric acid	3.98	4.6	Dns-putrescine	9.60	0.5
Dns-p-amino-hippuric acid	3.98	29	Dns-1,2 -diaminopropane	9.66	0.1
Dns-5-hydro xymelhyluricil	4.58	1.9	Dns-lyrosinamide	9.79	29
Dns-tryptophanamide	4.70	5.5	Dns-dopamine	10.08	140
Dris-isoguanine	4,75	-40L	Dns-cadaverine	10.08	0.08
Dns-5-aminopentanoic acid	4,79	1.6	Dns-histanine	10.19	0.4
Dns-sarcosine	4.81	7.2	Dns-3-mehoxy -tyramine	10.19	9.2
Dns-3-amino -isobutyrale	4.81	85	Dns-Tyr	10.28	321
Dns-2-aminobuly ric acid	4.91	17	Dns-cysteamine	10.44	⊲DL



From Pathways & Lists to Models & Biomarkers



Example LC-MS

Entity List 2 : Entities present

in Turning red control

2304 entities

Entity List 3 : Entities present

in Turning red treatment 1

2308 entities

- Blueberry Treatment 1 and 2
- Ripe vs Turning red



Example LC-MS



Treatment 2 has more of an effect on blueberries turning red

Serum and Adipose Tissue Amino Acid Metabolite Quantitation

Untargeted metabolite profiling



pp 3455–3466

Key Informatics Challenges in Metabolomics

- Spectra -> Lists
 - Data integrity and quality
 - Data alignment and normalization
 - Data reduction and classification
 - Assessment of significance
 - Metabolite identification/quantification
- Lists -> Pathways & Biomarkers
 - Pathway mapping and identification
 - Biological interpretation

Multi-Omics



Multi-Omics

F



Takahashi S, Saito K, Jia H, Kato H (2014) PLoS One 9:e91134

Multi-Omics



Beale DJ, Barratt R, Marlow DR, Dunn MS, Palombo EA, Morrison PD, Key C (2013) Biofouling 29:283–294

References

- Goodacre *et al* Analytica Chimica Acta 879 2015
 1-23.
- Julia *et al* Front Bioeng Biotechnol. 2015 **3** 1-20.
- Suizdak *et al* Nat Rev Mol Cell Bio 2016 13, 263-269.
- Suizdak *et al* Nat Rev Mol Cell Bio 2016 17, 451-459.
- Methods Mol Biol. 2014; **1198**: 333–353.

Metabolomic analysis of the cerebrospinal fluid reveals changes in phospholipase expression in the CNS of SIV-infected macaques



Fox et al, J. Clin Invest. 2008, 118, 2661.

Metabolomics

Table 3

Selected metabolites

Name	Molecular formula	<i>m/z</i> observed	<i>m/z</i> calculated	Relative error (ppm)	Retention time (min)	<i>P</i> value	Fold change
Carnitine	C7H15NO3	162.1124	162.1125	-0.62	2.9	0.010 (0.52)	9.1 (1.1)
Isovaleryl carnitine	C12H23NO4	246.1694	246.1699	-2.44	12.5	0.007 (0.35)	9.9 (1.2)
Octanoyl carnitine	C15H29NO4	288.2173	288.2169	1.04	23.9	0.005 (0.29)	3.7 (1.3)
Butyryl or isobutyryl carnitine	C11H21NO4	232.1548	232.1544	1.72	7.1	< 0.001 (0.36)	10.4 (1.8)
Palmitic acid (C16:0)	C16H32O2	257.2478	257.2475	1.17	47.2	0.011 (0.58)	35.5 (1.2)
Oleic acid (C18:1)	C18H34O2	283.2634	283.2632	0.71	39.5	0.029 (0.97)	2.5 (1.0)
Linoleic acid (C18:2)	C18H32O2	281.2472	281.2475	-1.07	40.7	0.004 (nd)	3.2 (nd)
Stearic acid (C18:0)	C18H36O2	285.2793	285.2788	1.75	50.9	0.005 (0.37)	14.4 (0.7)
Myristic acid (C14:0)	C14H28O2	229.2164	229.2162	0.87	43.0	0.009 (0.20)	5.5 (1.2)
LPC (16:0)	C24H50NO7P	496.3373	496.3398	-5.04	38.6	0.039 (0.59)	26.3 (1.2)
LPC (18:2)	C26H52NO7P	522.3542	522.3554	-2.30	39.5	0.042 (0.61)	27.5 (1.2)
LPC (18:0)	C26H54NO7P	524.3720	524.3711	1.72	43.5	0.038 (0.44)	73.9 (1.4)

Twelve metabolites were identified from the features that changed following SIV infection, based on mass accuracy and retention time. The observed and calculated *m/z* are given, with the differences shown in parts per million (ppm). The fold change reflects the average increase in samples taken after infection. *P* values (obtained using *t* test) and fold changes for animals that did not progress to encephalitis are shown in parentheses. nd, not determined.