

The Future of Metabolomics

David Wishart

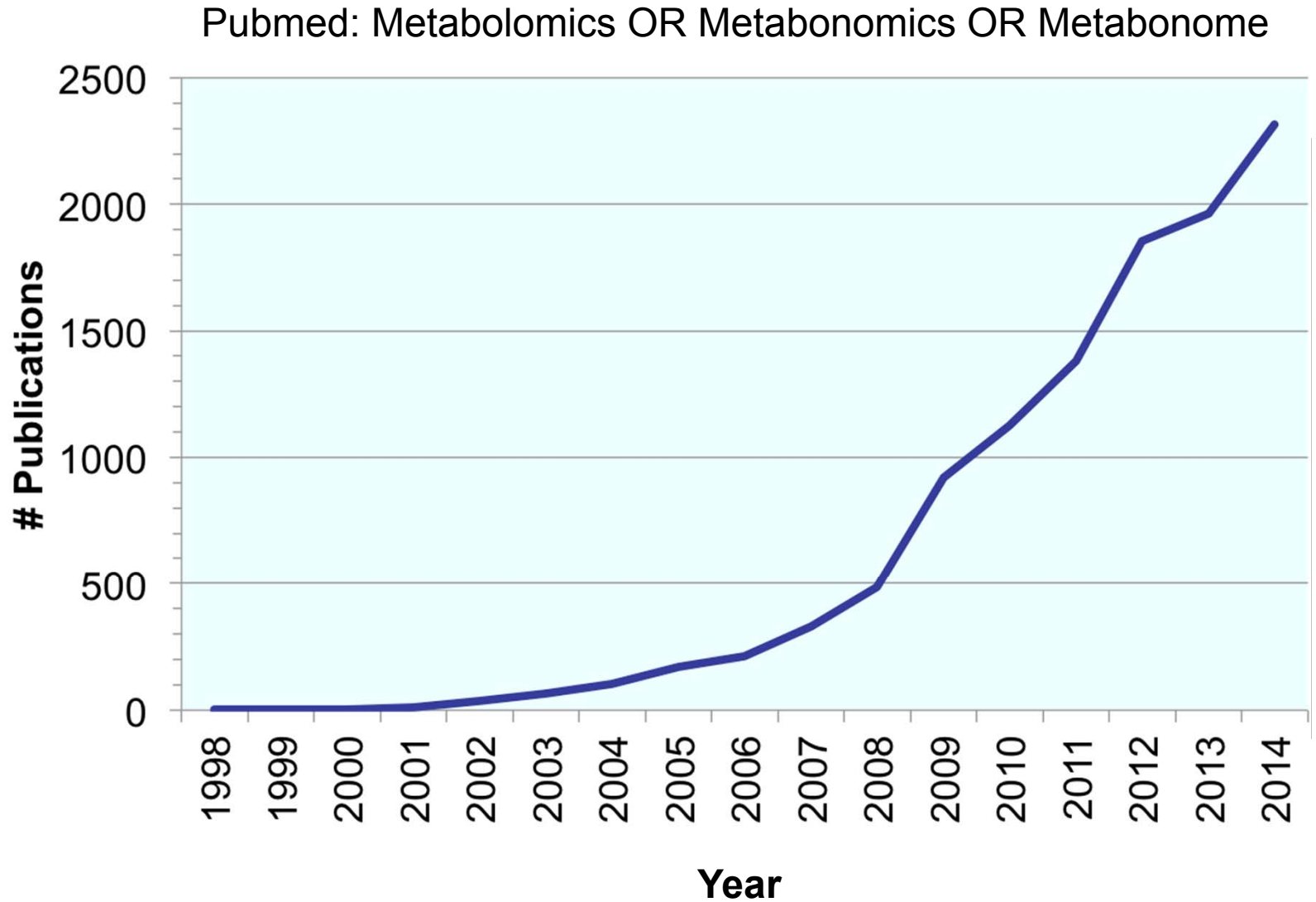
**University of Alberta, Edmonton, AB,
Canada**

Birmingham, Alabama June 18, 2015

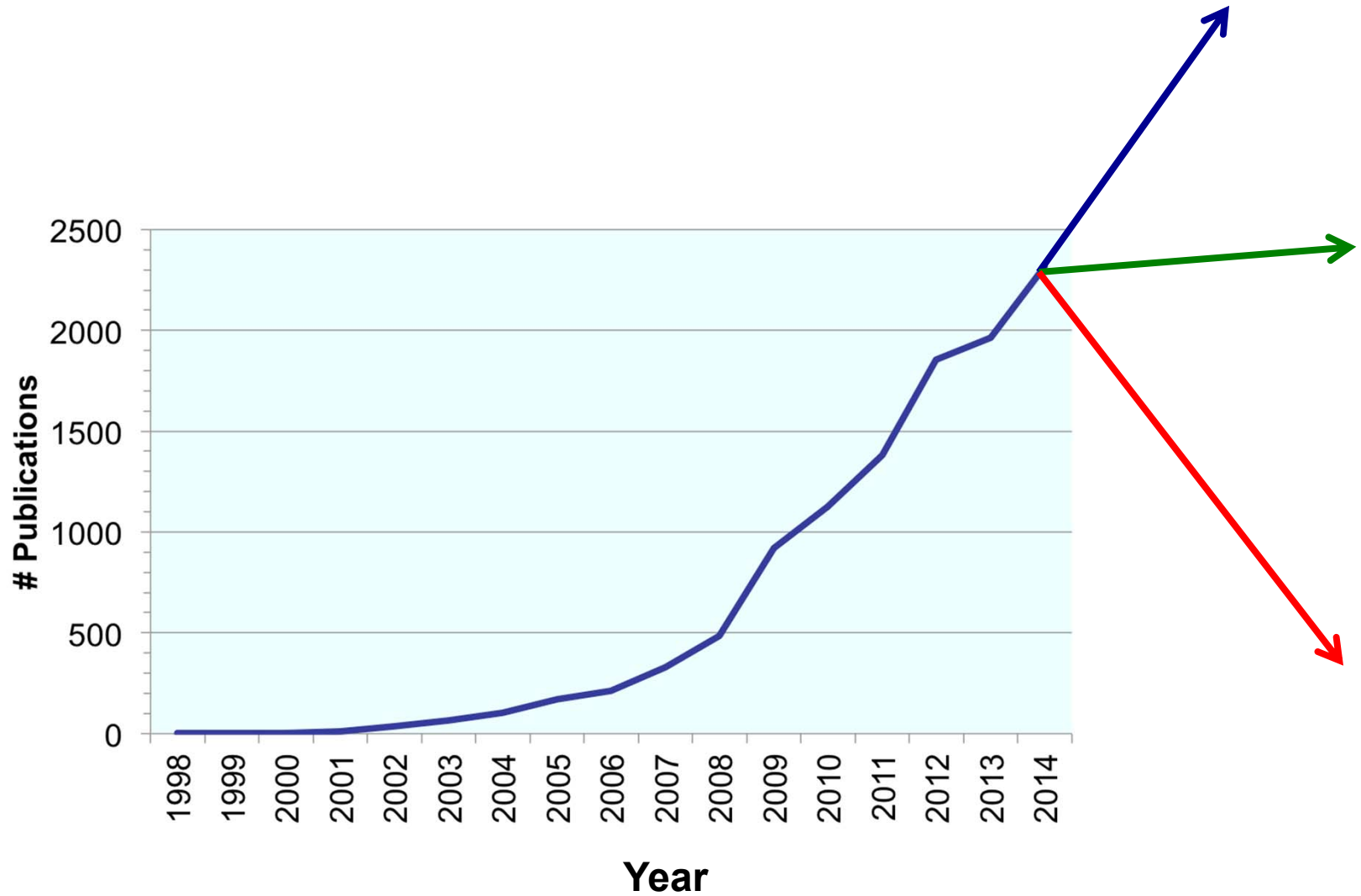
Why Small Molecules Matter

- **>99% of food flavours and aromas come from small molecules**
- **>90% of common clinical tests measure small molecules**
- **89% of all drugs are small molecules**
- **83% of the most common diseases are due to the effects of small molecules**
- **81% of all deaths in North America and Europe are due to the effects of small molecules**
- **>55% of drugs are derived via natural cmpds**

Metabolomics Is Growing



But Where Is It Going?



Key Bottlenecks in Metabolomics

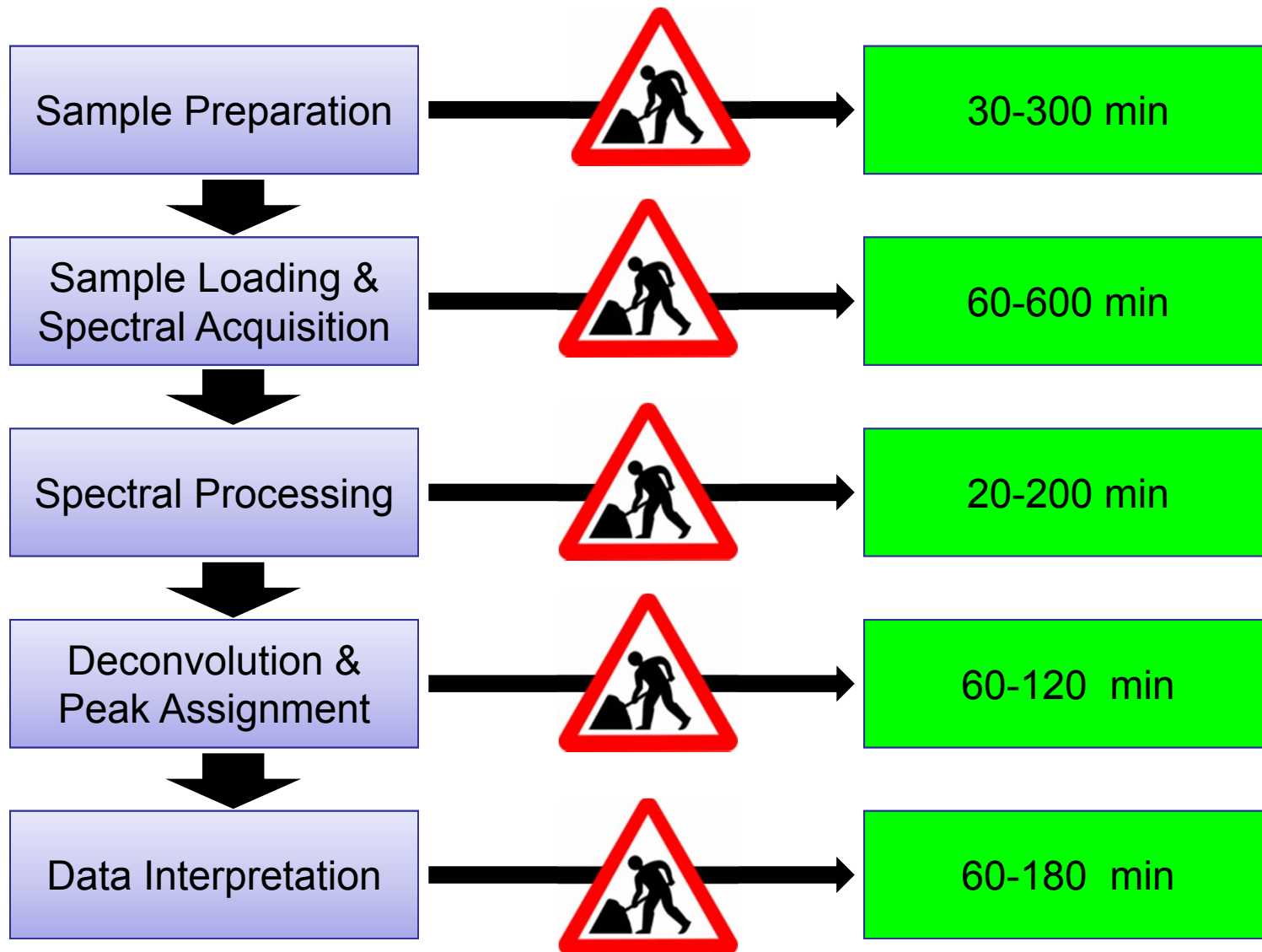
- **Lack of automation**
- **Incomplete metabolome coverage**
- **Expensive/large equipment**
- **Lack of quantification**
- **Inability to translate findings to the clinic**
- **Making metabolomics matter to drug companies**



Key Trends in Metabolomics

- **Automated metabolomics**
- **Expanding metabolome coverage**
- **Making metabolomics portable**
- **Quantify, quantify, quantify...**
- **Moving metabolomics from the lab to the clinic**
- **Moving metabolomics (back) into drug development and discovery**

Metabolomics Workflow



Automated Metabolomics

Targeted Analysis

Compound	Value	Unit	Official Reference	Wine Profiling
2,3-butanediol	326	mg/L		
3-methyl butanol	111	mg/L		
IBP	<5	mg/L		
acetaldehyde	<5	mg/L		
acetic acid	418	mg/L		
acetone	9	mg/L		
alcohols	79	mg/L		
benzoic acid	<5	mg/L	0 mg/L ¹⁾	
lactic acid	88	mg/L		
citric acid	<200	mg/L	1000 mg/L ¹⁾	
ethanol	95.6	g/L		
formic acid	<5	mg/L		
fructose	3.8	g/L		
gluconic acid	<5	mg/L		
glycerol	5.1	g/L		
lipoic acid	1.0	g/L		
malic acid	1.6	g/L		
methanol	75	mg/L	250 mg/L ¹⁾	
phthalic acid	54	mg/L		
sorbic acid	<5	mg/L	200 mg/L ¹⁾	
succinic acid	285	mg/L		
tartronic acid	<5.2	g/L		
tartaric acid	1.8	g/L		
triglycerides	8	mg/L		

Source for Reference Values
1) OIV Reference
2) Maximum Value in wine



Bruker – Automated NMR

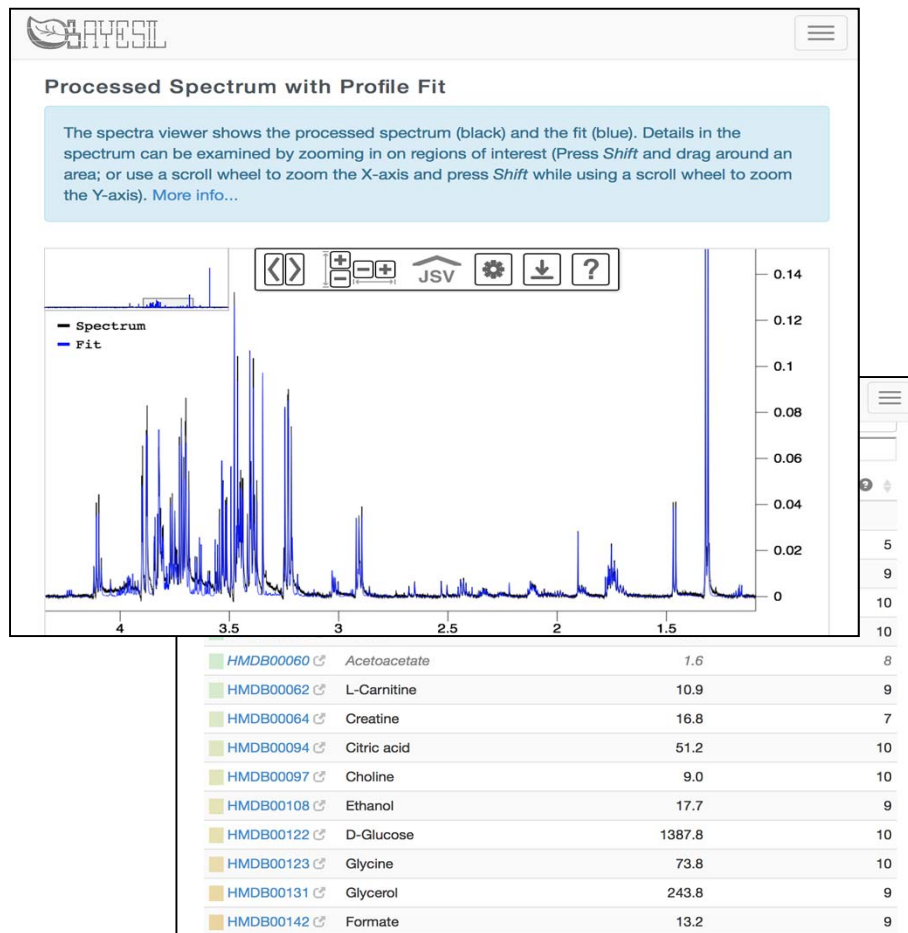


Sciex – Lipidyzer



Biocrates – Automated MS

Bayesil (Automated NMR)



<http://bayesil.ca>

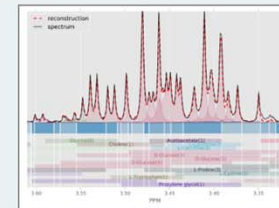
- Uses probabilistic graphical models (PGM) – similar to HMMs
- Fits shift & peak intensity similar to the way humans perform fitting and pattern finding
- Requires prior knowledge of probable biofluid composition
- Fully automated phasing, referencing, water removal, baseline correction, peak convolution, identification and quantification
- Free web server

Bayesil in Operation



Welcome to Bayesil

Bayesil is a web system that automatically identifies and quantifies metabolites using 1D ^1H NMR spectra of ultra-filtered plasma, serum or cerebrospinal fluid. The NMR spectra must be collected in a standardized fashion (see [How To Collect NMR Spectra for Bayesil](#)) for Bayesil to perform optimally. Bayesil first performs all spectral processing steps, including Fourier transformation, phasing, solvent filtering, chemical shift referencing, baseline correction and reference line shape convolution automatically. It then deconvolutes the resulting NMR spectrum using a reference spectral library, which here contains the signatures of more than 60 metabolites (see [here](#) for a list). This deconvolution process determines both the identity and quantity of the compounds in the biofluid mixture. Extensive testing shows that Bayesil meets or exceeds the performance of highly trained human experts.



Citing Bayesil:

Ravanbakhsh S, Liu P, Bjordahl TC, Mandal R, Grant JR, Wilson M, Eisner R, Sinelnikov I, Hu X, Luchinat C, Greiner R, Wishart DS. (2015) Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics. PLoS ONE 10(5): e0124219.

C

Bayesil Spectral Analysis

Instructions

To analyze a 1D ^1H NMR spectrum with Bayesil you must provide information on the biofluid being analyzed, the concentration of the reference standard, the spectrometer frequency and the 1D NMR spectral file.

Run one of our examples:

Example 1

Biological Serum
Varian 500 MHz

Example 2

Biological CSF
Varian 500 MHz

Example 3

Biological Serum
Bruker 600 MHz

ec

S. Ravanbakhsh, et al. (2015) PLoS One 10(5): e0124219

GC-AutoFit (Automated GC-MS)

GCMS auto-profiling

Home Check Result Check Job ID Instructions Contact Us

Welcome to GC-AutoFit

GC-AutoFit is a web application that automatically identifies and quantifies metabolites using Gas Chromatography Mass Spectrometry (GC-MS) spectra. For optimal GC-AutoFit performance, the query GC-MS spectra should be prepared according to the [instructions here](#). GC-AutoFit currently accepts .CDF and .mzXML file formats. It uses alkane standards to calculate the retention index (RI) of each peak in the sample. The extracted EI-MS spectra from each peak, along with the RIs, are then compared to reference spectra (RIs and EI-MS) in the specified library to identify and quantify the compounds. The inclusion of blank spectra is optional, however, it is useful for removing noise effects from the query spectra. Extensive testing shows that GC-AutoFit meets or exceeds the performance of highly trained human experts.

Upload Spectrum

.CDF or .mzXML (suffix required) format is acceptable. Alternatively, a single .zip file containing all spectrum files can be used. If you wish to use your own library, it should be in .csv (suffix required) format. The total number of files cannot exceed 33.

Alkane Standards file (Required): No file selected.

Blank files (Optional): No file selected.

Sample files (Required): No file selected.

OR

Zip file: No file selected.

Database (Required): Internal library Your own library

Submit Reset

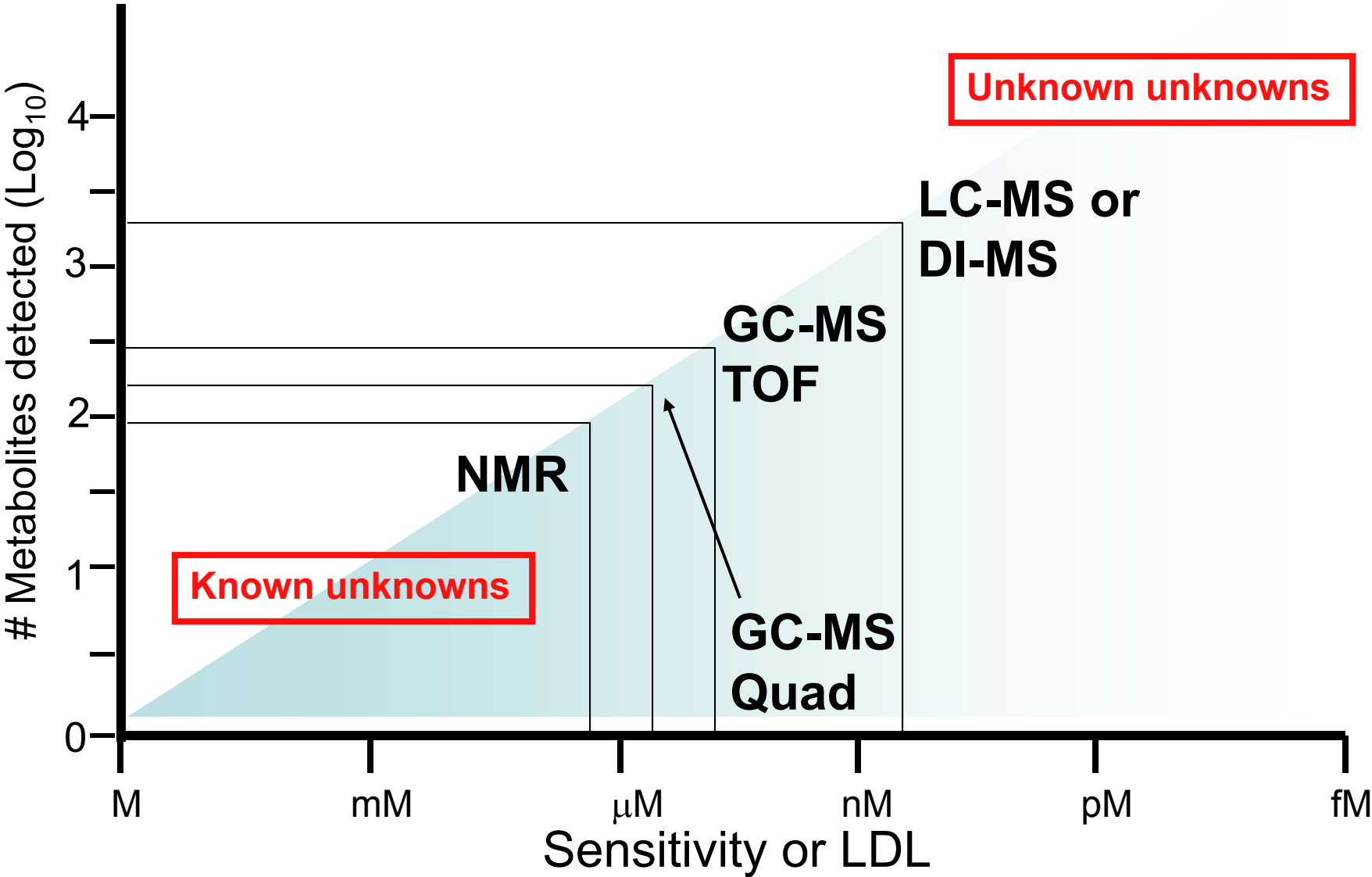
<http://gcms.wishartlab.com>

- Requires 3 spectra (sample, blank, alkane standards)
- Performs auto-alignment, peak ID, peak integration and concentration calculation
- Accepts NetCDF or mzXML files
- 60 sec per spectrum
- 45-70 cmpds ID' d and quantified, 96% accuracy
- Optimized for blood, urine, saliva and CSF
- Still requires careful sample preparation & derivatization

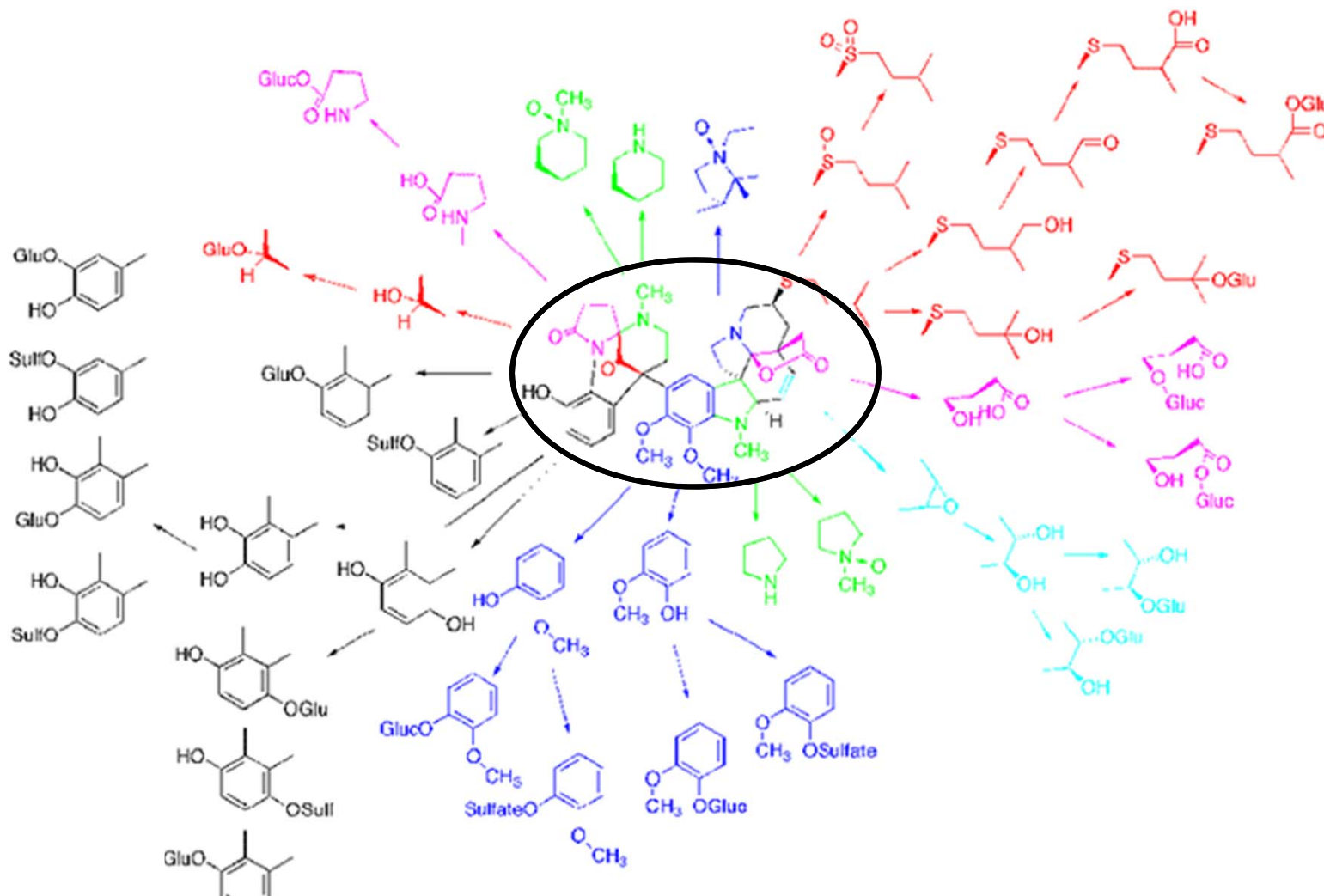
Key Trends in Metabolomics

- Automated metabolomics
- Expanding metabolome coverage
- Making metabolomics portable
- Quantify, quantify, quantify...
- Moving metabolomics from the lab to the clinic
- Moving metabolomics (back) into drug development and discovery

Technology & Sensitivity

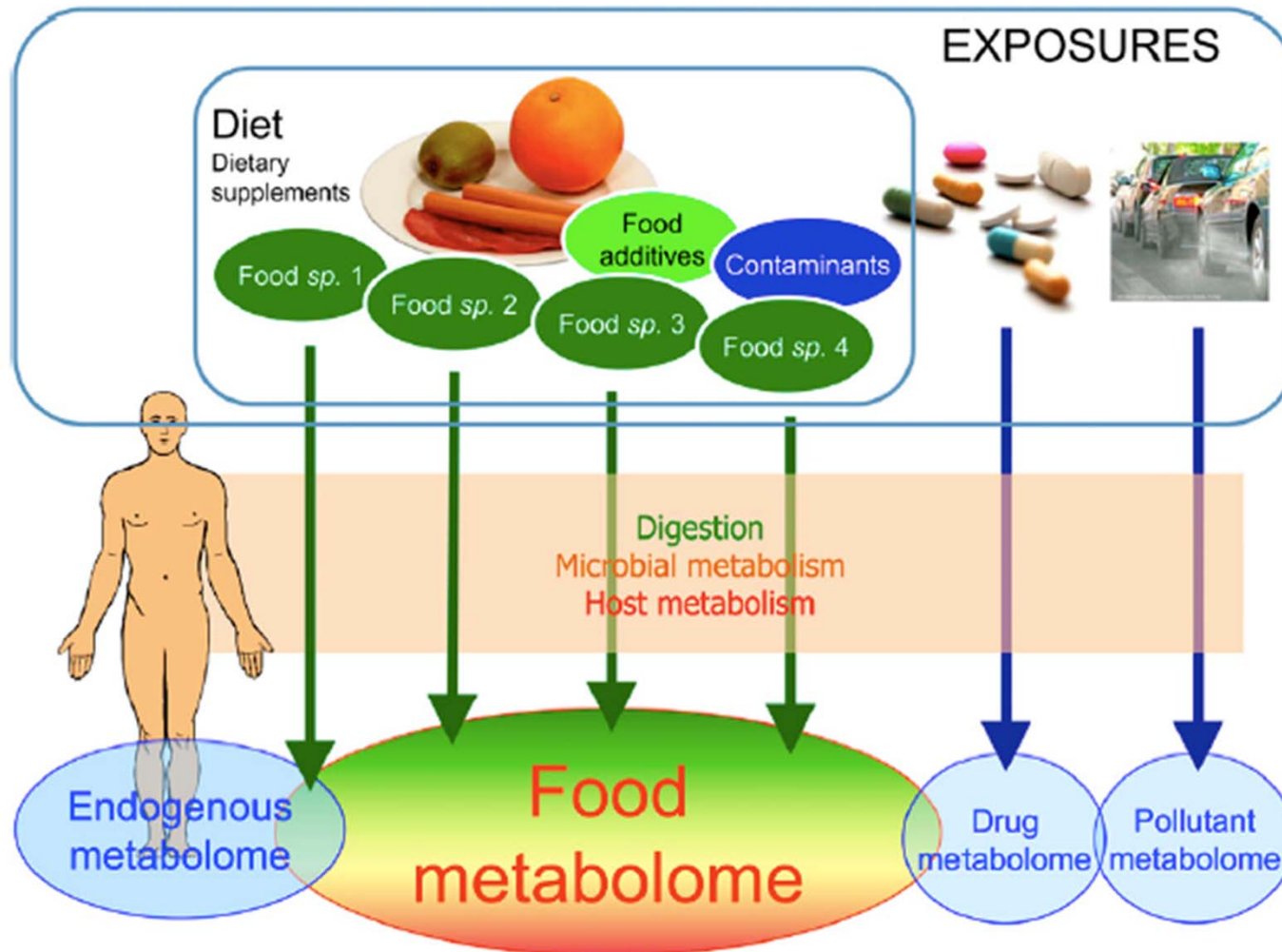


What Are The Unknown Unknowns?



Metabolites of Metabolites

The Food Metabolome



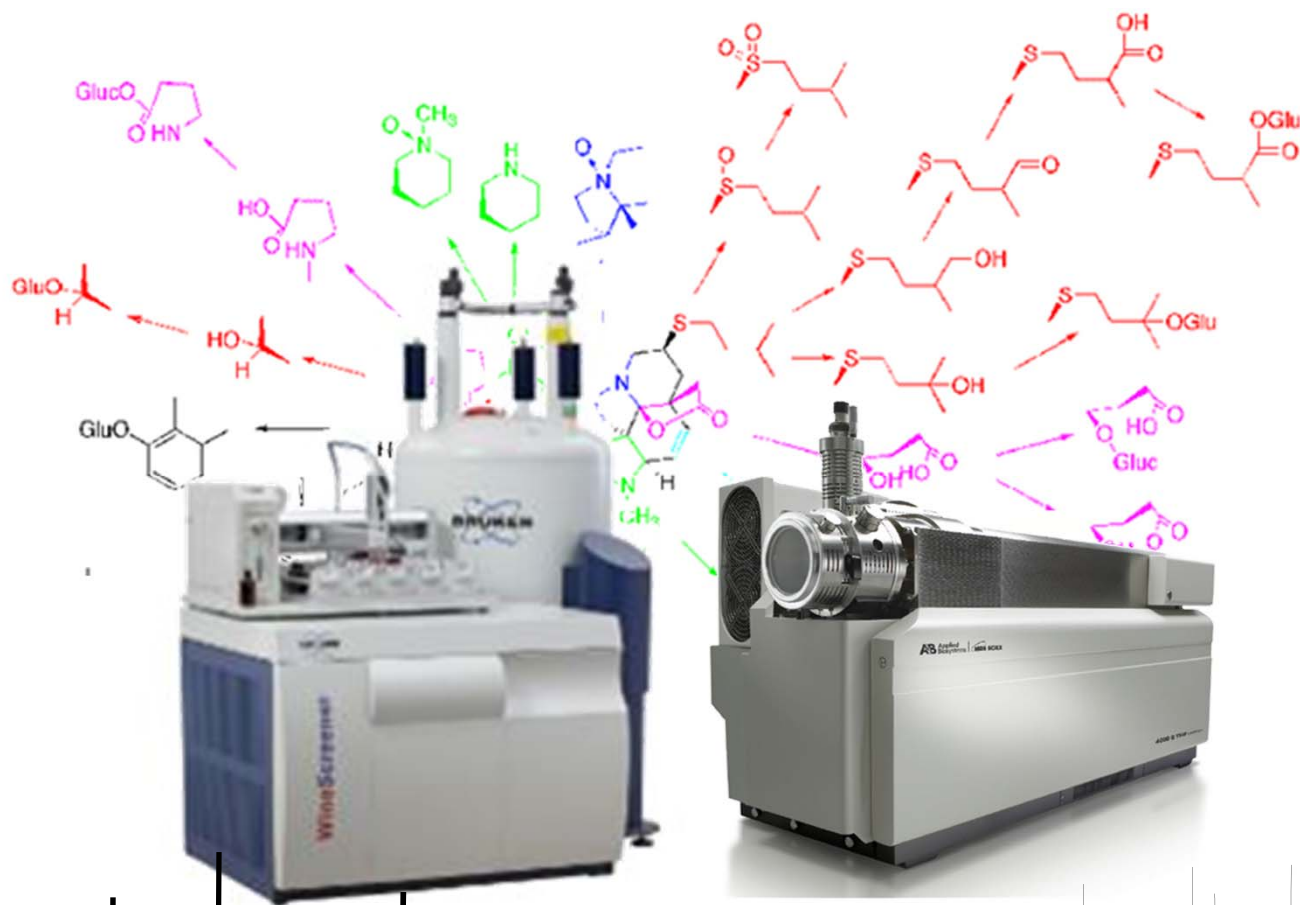
Scalbert A. et al. (2014) Am. J. Clinical Nutr. 99(6):1286

Systematic Spectral Collection

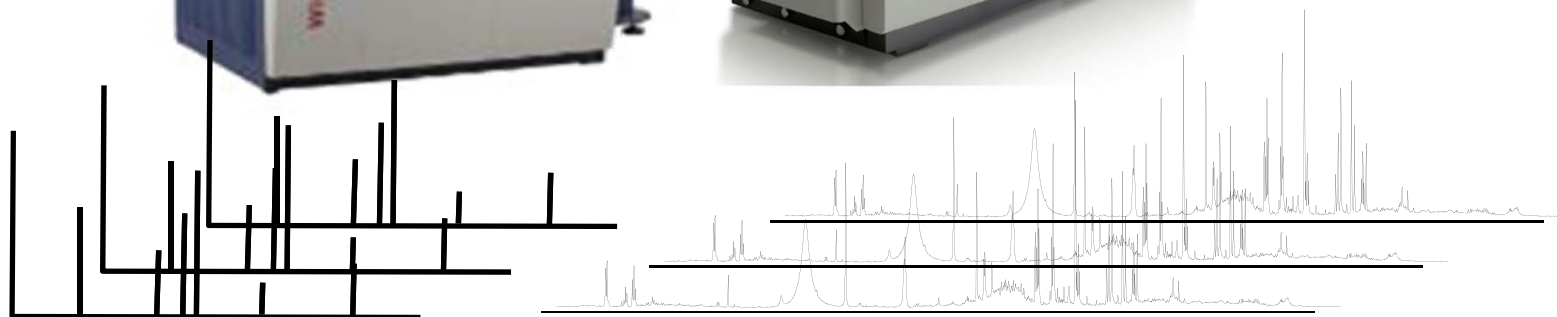
100,000+
Compounds

\$\$

800,000+
NMR, MS/MS
Spectra



\$\$



Systematic Spectral Prediction

The screenshot displays the CFM-ID web application interface. At the top, there is a navigation bar with 'CFM-ID' and links for 'Utilities', 'Help', 'Data', and 'Contact Us'. Below this, a welcome message reads 'Welcome to CFM-ID!'. The main content area is divided into several sections:

- CFM-ID Description:** A text box explaining that CFM-ID provides a method for accurately and efficiently identifying metabolites in spectra generated by electrospray tandem mass spectrometry (ESI-MS/MS). It uses Competitive Fragmentation Modeling to produce a probabilistic generative model for the MS/MS fragmentation process and machine learning techniques to adapt the model parameters from data.
- Workflow Steps:** A vertical list of steps: 'Spectra Prediction', 'Peak Assignment', and 'Compound Identification'. Each step has a brief description of its function.
- CFM-ID: Results:** A section showing the results of a prediction. It includes three mass spectra plots: 'Low Energy Input MsMs Spectrum (10V), M+H', 'Medium Energy Input MsMs Spectrum (20V), M+H', and 'High Energy Input MsMs Spectrum (40V), M+H'. The first plot shows a base peak at m/z 296. Below the spectra is a 'Candidate Rankings' table.

Rank	Score	Structure	ID	Name	Chemical Formula	Mass	Compare
1	0.28571429		HMDB29101	Tyrosyl-Aspartate	C13H16N2O6	296.100836254	<input type="button" value="Current"/>
2	0.18047619		HMDB28765	Aspartyl-Tyrosine	C13H16N2O6	296.100836254	<input type="button" value="Compare"/>
3	0.076823077		HMDB11665	DHAP(B:0)	C11H21O7P	296.102489538	<input type="button" value="Compare"/>

<http://cfmid.wishartlab.com>

- Predicts MS/MS spectra from known compounds via advanced machine learning techniques
- 50% more accurate than other systems
- Matches predicted MS/MS spectra (from HMDB, KEGG or user choice) to input MS/MS spectra
- Permits rapid compound ID from MS/MS spectra

Predicting Metabolites of Metabolites

The screenshot shows the MyCompoundID web application interface. The browser address bar displays the URL: <http://beiseker.cs.ualberta.ca:8080/mycompoundid/single.jsp>. The page title is "Search - MyCompoundID". The interface includes a search bar with the text "131.094 Da (Batch Mode)" and a "Submit Query" button. Below the search bar, there are radio buttons for "No reaction", "1 reaction", and "2 reactions". There are also radio buttons for "Neutral" and "Neutral or ion:" with sub-options for various ionization states. A "Mass Tolerance" section has radio buttons for "In Da (default: ± 0.005 Da)" and "In ppm (default: ± 5 ppm)".

11	-H ₂ O	-18.010565	loss of water
12	+H ₂ O	18.010565	addition of water
13	-CO	-27.994915	loss of CO
14	+CO	27.994915	addition of CO
15	-C ₂ H ₄	-28.031300	loss of C ₂ H ₄
16	+C ₂ H ₄	28.031300	addition of C ₂ H ₄
17	-C ₂ H ₂ O	-42.010565	Deacetylation
18	+C ₂ H ₂ O	42.010565	Acetylation
19	-CO ₂	-43.989830	loss of CO ₂
20	+CO ₂	43.989830	addition of CO ₂
21	SO ₃ H->SH	-47.984745	Sulfonic acid to Thiol
22	SH->SO ₃ H	47.984745	Thiol to Sulfonic acid
23	-C ₂ H ₃ NO	-57.021464	loss of glycine
24	+C ₂ H ₃ NO	57.021464	glycine conjugation
25	-SO ₃	-79.956817	loss of sulfate
26	+SO ₃	79.956817	sulfate conjugation
27	-HPO ₃	-79.966333	loss of Phosphate

- Calculates the MW of metabolic transformations (+adducts, neutral loss fragments) from HMDB “parent” metabolites
- 375,809 compounds from one metabolic reaction and 10,583,901 from two reactions
- Number of putative compound hits (via mass matching) for MS-based metabolomic experiments increases 4-5X

<http://mycompoundid.org>

Key Trends in Metabolomics

- Automated metabolomics
- Expanding metabolome coverage
- Making metabolomics portable
- Quantify, quantify, quantify...
- Moving metabolomics from the lab to the clinic
- Moving metabolomics (back) into drug development and discovery

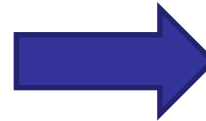
Personalized Medical Monitoring Devices



Democratizing Metabolomics



\$10 million instrument, \$200/test



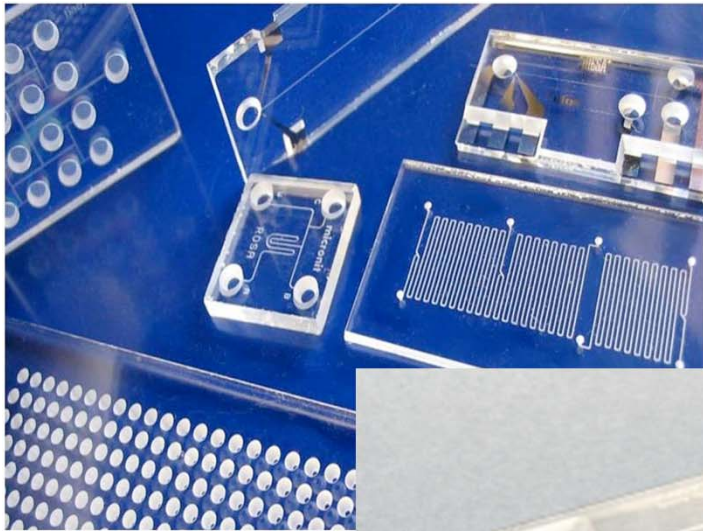
\$1000 instrument, \$2/test

Not As Absurd As You Think

QUALCOMM
TRICORDER X PRIZE



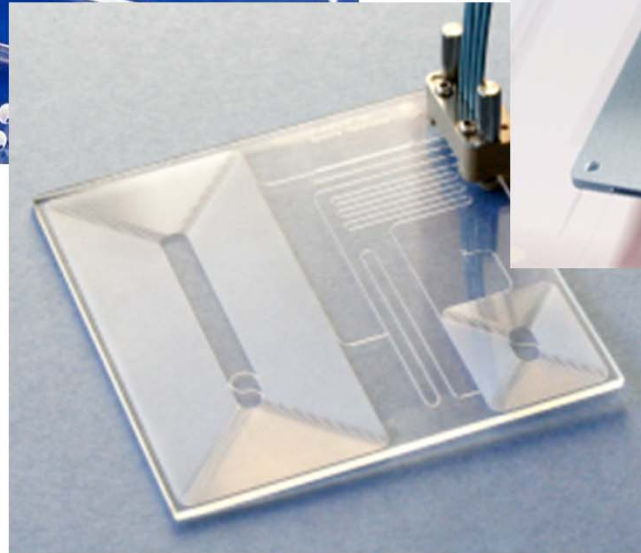
Miniaturization via Microfluidics & Nanotech



CE on a chip

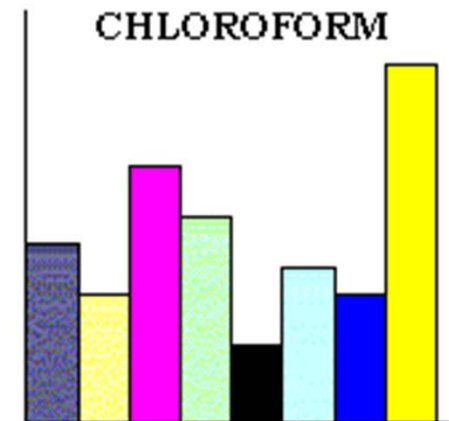
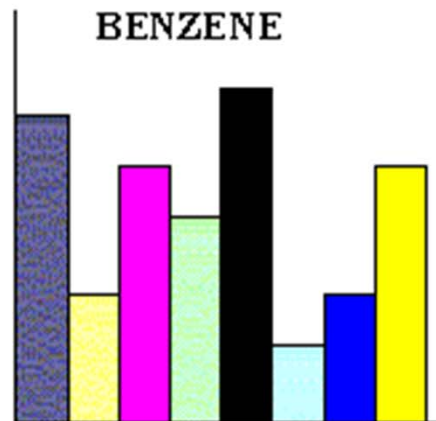
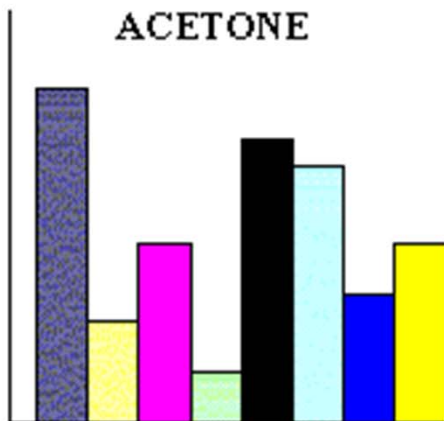
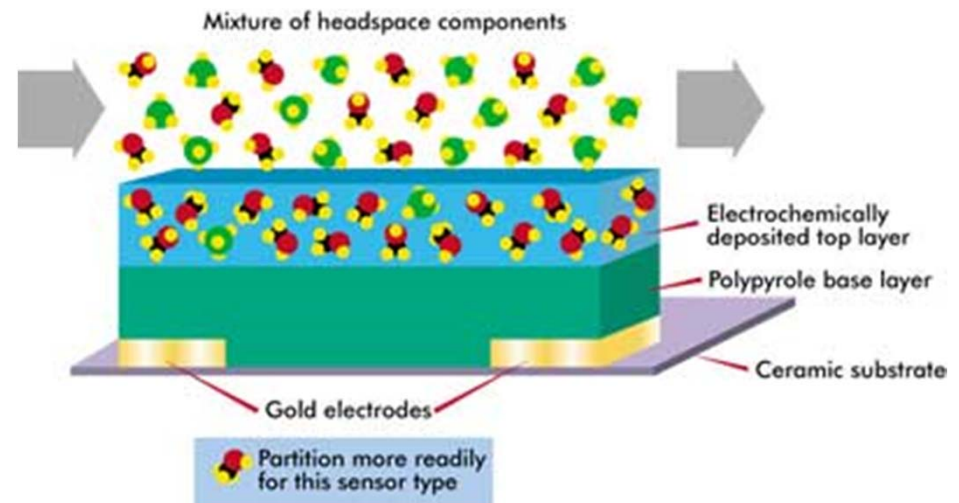


HPLC on a chip

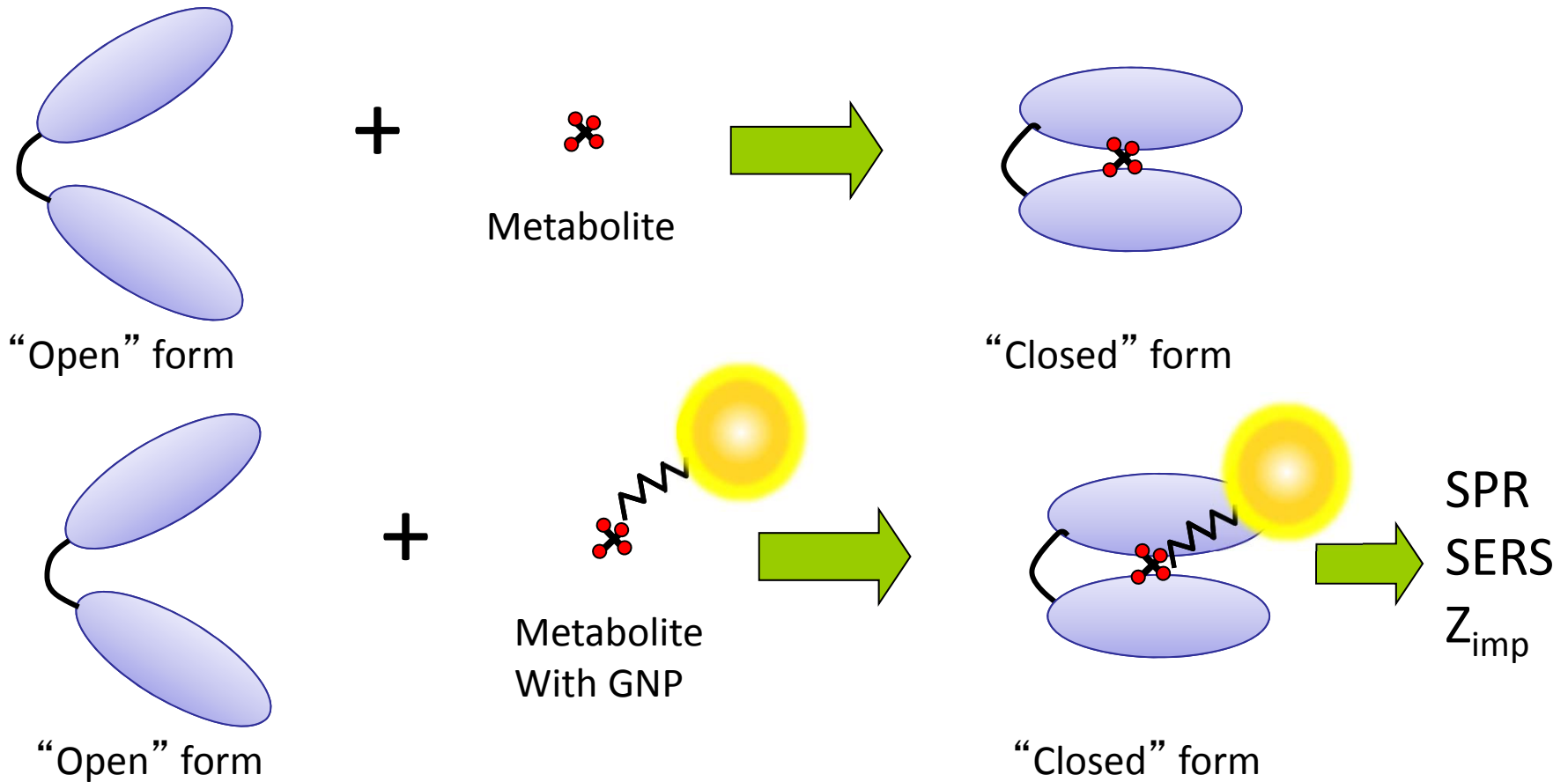


GC on a chip

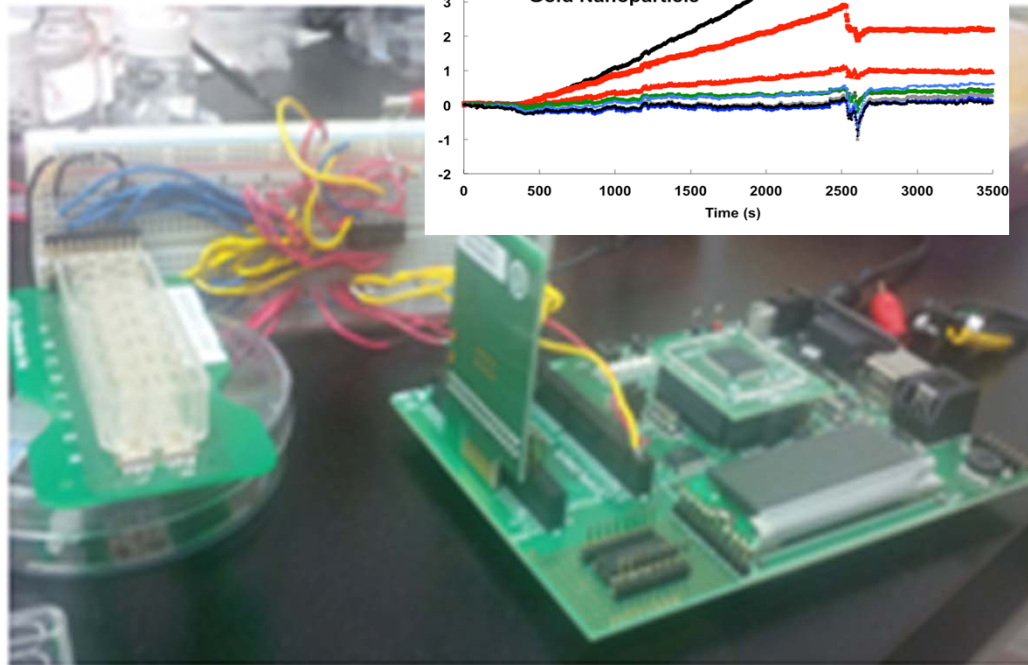
E-Nose for Volatile Metabolites



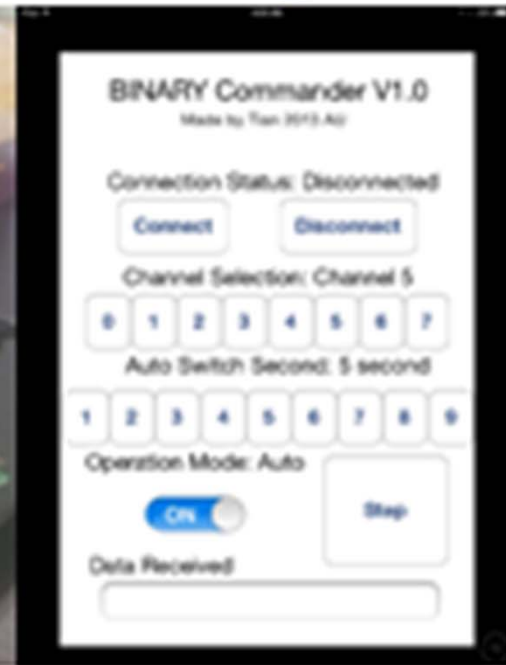
Protein and Aptamer-Mediated Metabolite Sensing



Working Prototype: Impedance-Based Metabolite Sensor



iPhone App



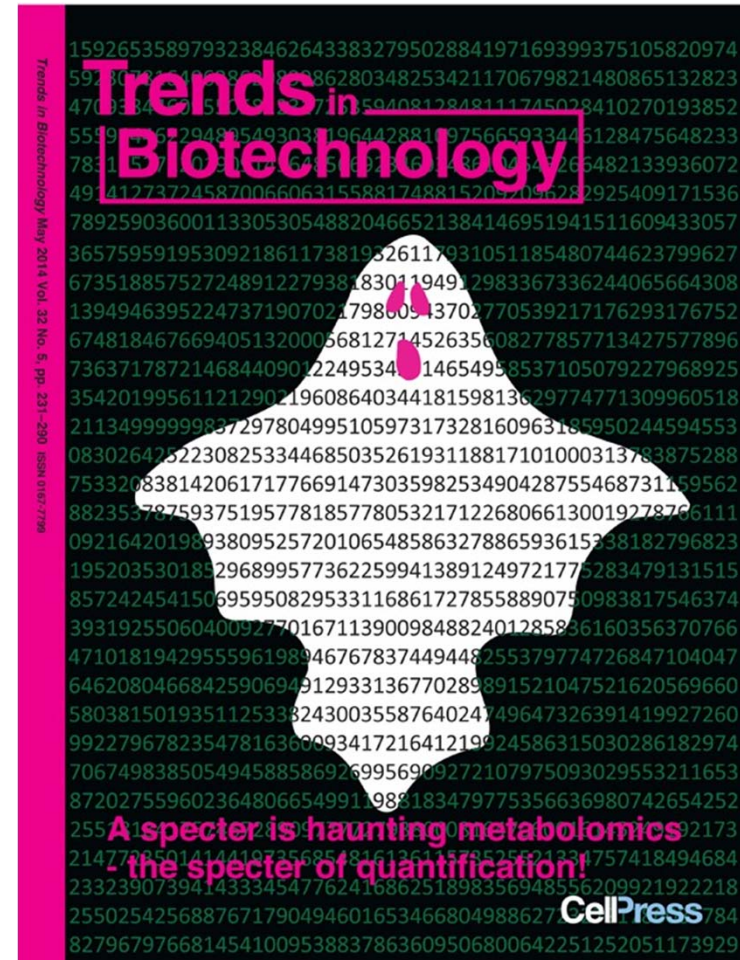
Developed by Dr. Jie Chen, University of Alberta

Key Trends in Metabolomics

- Automated metabolomics
- Expanding metabolome coverage
- Making metabolomics portable
- **Quantify, quantify, quantify...**
- Moving metabolomics from the lab to the clinic
- Moving metabolomics (back) into drug development and discovery

Quantification & Metabolomics

- >90% of published metabolomics studies are semi-quantitative (relative peak areas, intensities)
- <10% of published metabolomics studies use absolute quantification
- The field **MUST** become more quantitative if findings are to be translated to practical applications



Quantitative Metabolomics (Commercial)

DISCOVER MORE

CHENOMX NMR SUITE is an integrated suite of tools allowing you to identify and quantify compounds in NMR spectra.

- Identify and quantify hundreds of small molecules
- Analyze virtually any biofluid
- Multi-platform support
- Intuitive user interface

Targeted Analysis

Compound	Value	Unit	Flag	Optical Reference	Peak	Value	View Plotting
2,3-butanediol	326	mg/L					
3-methyl butanol	111	mg/L					
HOAc	<5	mg/L					
acetabutylic	<5	mg/L					
acetic acid	448	mg/L					
alanine	9	mg/L					
alanine	79	mg/L					
amino acid	<5	mg/L					
aspartic acid	69	mg/L					
citric acid	205	mg/L					
ethanol	963	g/L					
formic acid	<5	mg/L					
fructose	3.8	g/L					
glucose	<5	mg/L					
glucose	3.3	g/L					
glucose	8.1	g/L					
lactic acid	1.0	g/L					
malic acid	28	g/L					
malic acid	75	mg/L					
malonic acid	54	mg/L					
malonic acid	<5	mg/L					
malonic acid	205	mg/L					
malonic acid	<5	mg/L					
malonic acid	<0.2	g/L					
malonic acid	3.8	g/L					
triglyceride							

Software

Features

What's New

Download

Buy It Now

Compound List

Try it Now!

CHENOMX NMR SUITE 7.7

Compound Name	Concentration (ppm)
Alanine	0.833
Aspartic acid	1.111
Citric acid	1.333
Glucose	1.555
Lactic acid	1.777
Malic acid	2.000
Malonic acid	2.222
Malonic acid	2.444
Malonic acid	2.666
Malonic acid	2.888
Malonic acid	3.111
Malonic acid	3.333
Malonic acid	3.555
Malonic acid	3.777
Malonic acid	4.000
Malonic acid	4.222
Malonic acid	4.444
Malonic acid	4.666
Malonic acid	4.888
Malonic acid	5.111
Malonic acid	5.333
Malonic acid	5.555
Malonic acid	5.777
Malonic acid	6.000
Malonic acid	6.222
Malonic acid	6.444
Malonic acid	6.666
Malonic acid	6.888
Malonic acid	7.111
Malonic acid	7.333
Malonic acid	7.555
Malonic acid	7.777
Malonic acid	8.000
Malonic acid	8.222
Malonic acid	8.444
Malonic acid	8.666
Malonic acid	8.888
Malonic acid	9.111
Malonic acid	9.333
Malonic acid	9.555
Malonic acid	9.777
Malonic acid	10.000

Chenomx – Automated NMR

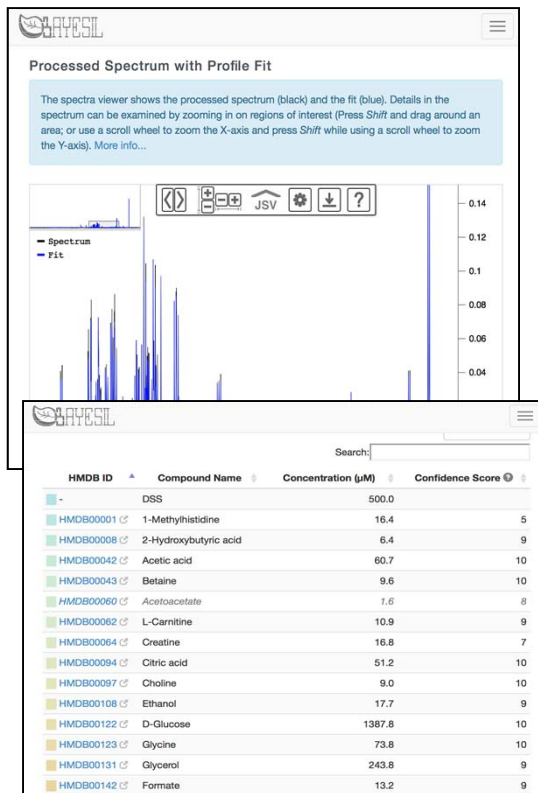


Bruker – Automated NMR

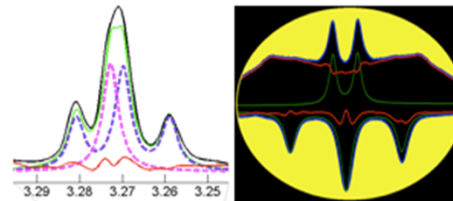
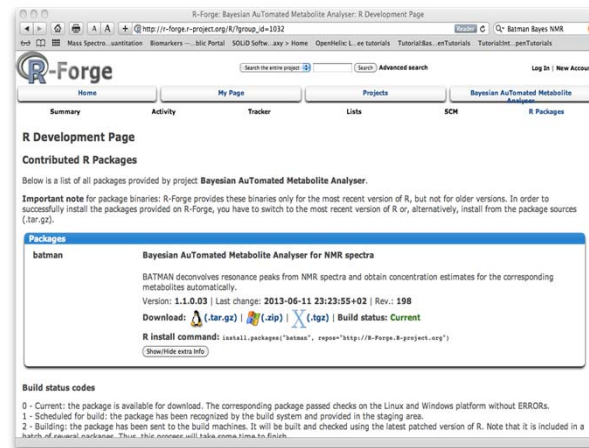


Biocrates – Automated MS

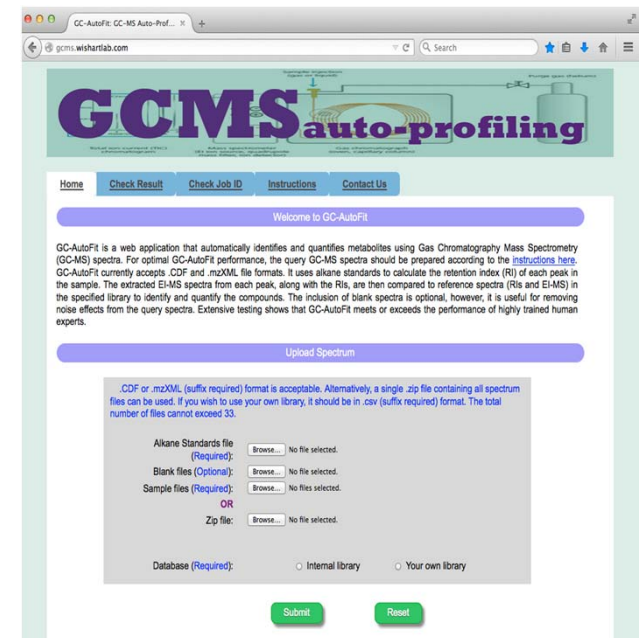
Quantitative Metabolomics (Academic)



Bayesil



Batman



GC-Autofit

Some Impressive Results...

Human Biofluid Omics “Records” for Absolute Quantification

	Metabolomics	Proteomics	Genomics (Transcripts)
Serum/Plasma	288 Identified & Quantified ¹	73 Identified & Quantified ⁴	0
CSF	172 Identified & Quantified ²	130 Identified & Quantified ⁵	0
Urine	378 Identified & Quantified ³	63 Identified & Quantified ⁶	0

1. Psychogios N. et al. (2011) PLoS One 6(2): e16957
2. Mandal R. et al. (2012) Genome Med.;4(4):38.
3. Bouatra S. et al. (2013) PLoS One 8(9): e73076
4. MRM Proteomics Inc. (Victoria BC) reported in 2014
5. Percy AJ. et al. (2014) J. Proteome Res. (ePub Jun 9)
6. Chen YT. et al. (2012) J. Proteomics 75(12):3529

Key Trends in Metabolomics

- Automated metabolomics
- Expanding metabolome coverage
- Making metabolomics portable
- Quantify, quantify, quantify...
- **Moving metabolomics from the lab to the clinic**
- Moving metabolomics (back) into drug development and discovery

Some Grim Statistics

- **Since 1970 > 700,000 biomarker papers published in PubMed**
- **Since 1970 <250 biomarkers have been approved for clinical use**
- **No markers approved (yet) using proteomics methods (lots use ELISA)**
- **5 biomarker tests approved using transcriptomics or gene chips**

But Did You Know... Almost Everyone <25 Has Had A Metabolomic Test?



Newborn Screening



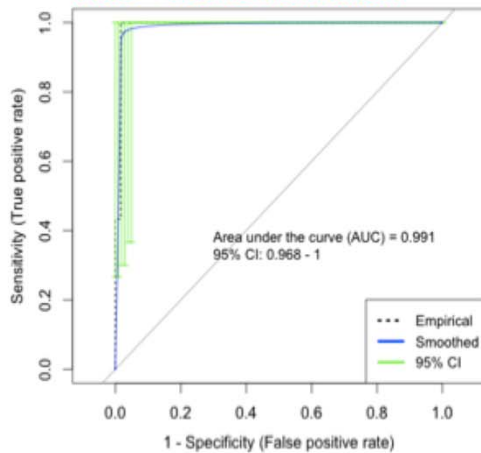
“Omics” Testing

- Number of “approved” tests arising from **Metabolomics/Clinical Chem.** – **195**
- Number of “approved” tests arising from or using **Genomics** – **100-110**
- Number of “approved” single **Protein** tests (ELISA) – **60**
- Number of “approved” tests arising from or using **Transcriptomics** – **5**
- Number of “approved” tests arising from or using **Proteomics** - **0**

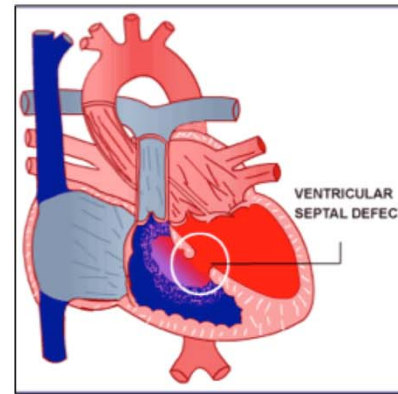
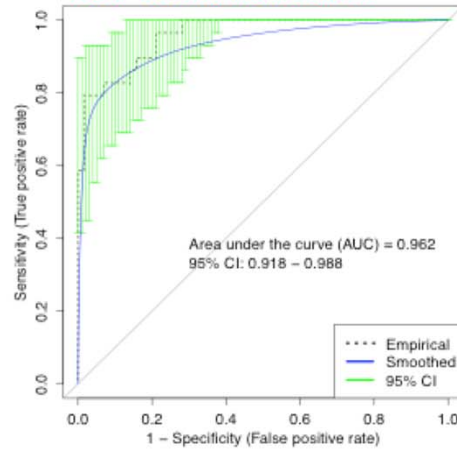
How Does Metabolomics Do? (Prediction & Diagnosis)

Predicting Diseases

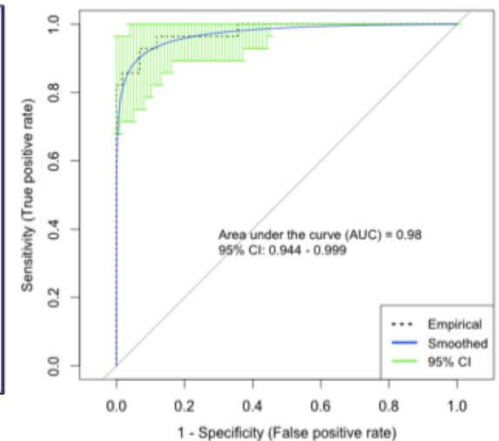
Early Preeclampsia
First trimester maternal serum
AUC = 0.99 (2 metabolites)



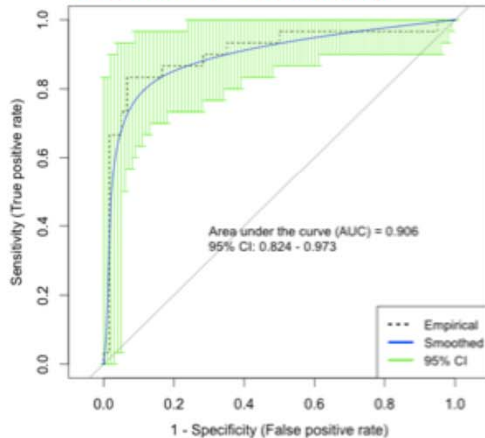
Late Preeclampsia
First trimester maternal serum
AUC=0.96 (8 metabolites)



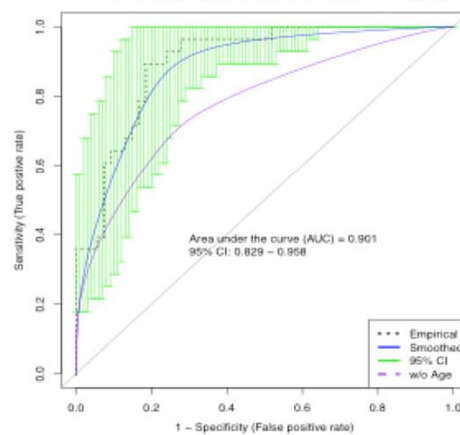
Congenital Heart Defects (CHD)
Maternal Serum
AUC=0.98 (3 metabolites)



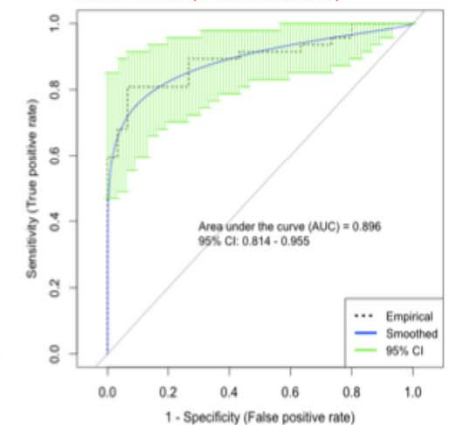
Trisomy 18
First trimester maternal serum
AUC=0.91 (7 metabolites)



Trisomy 21
First trimester maternal serum
AUC=0.90 (3 metabolites + Age)

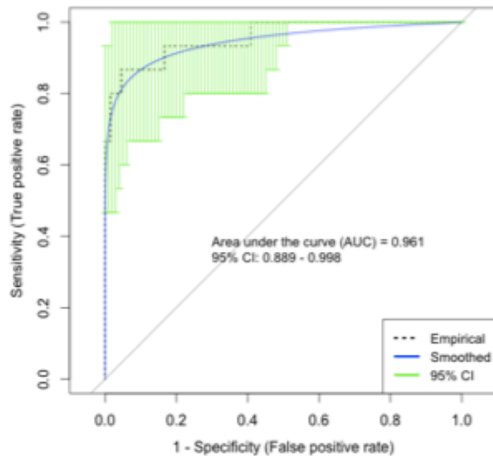


Cancer Cachexia
Adult Urine samples
AUC=0.90 (4 metabolites)

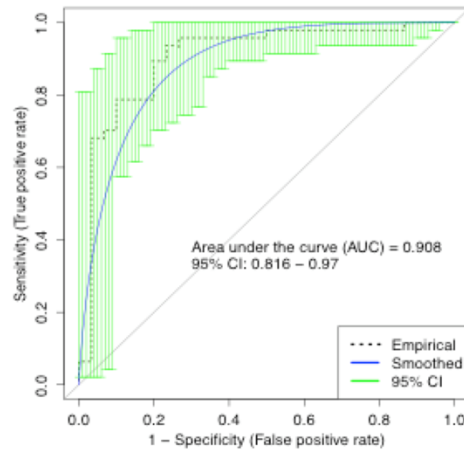


Diagnosing Diseases

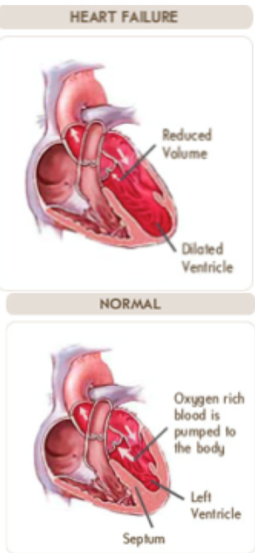
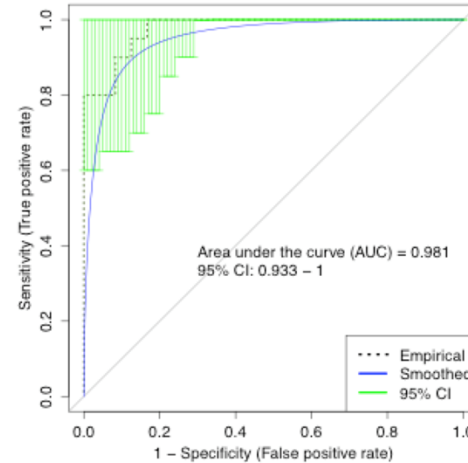
Adult kidney transplant rejection
Adult Urine samples
AUC=0.96 (9 metabolites)



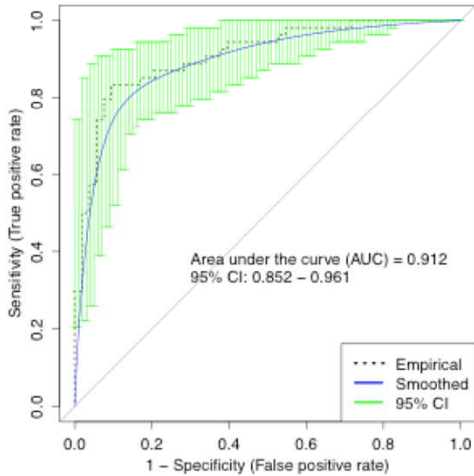
Pediatric kidney transplant rejection
Pediatric Urine samples
AUC=0.91 (4 metabolites)



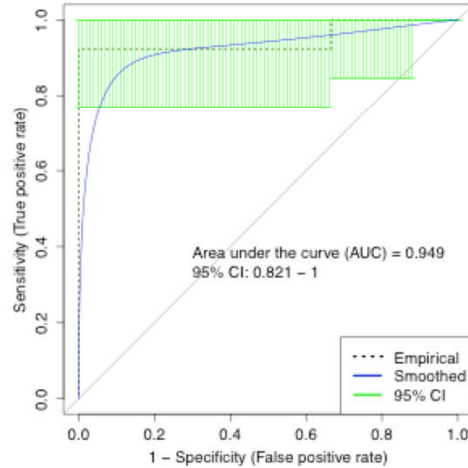
Systolic Heart Failure vs. Diastolic HF
Adult Serum samples
AUC=0.98 (4 metabolites)



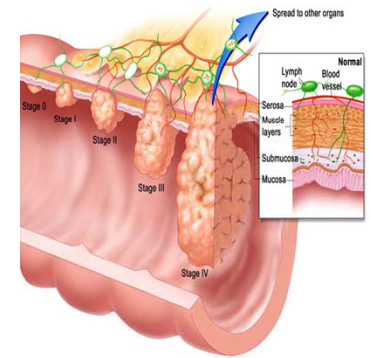
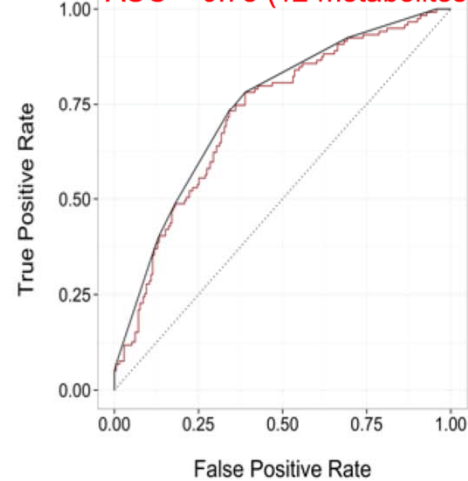
Chronic Fatigue Syndrome
Adult Serum samples
AUC=0.91 (6 metabolites)



Eosinophilic Esophagitis (EoE)
Pediatric Urine Samples
AUC=0.95 (3 metabolites)



Colonic Polyps
Adult Urine Samples
AUC = 0.78 (12 metabolites)

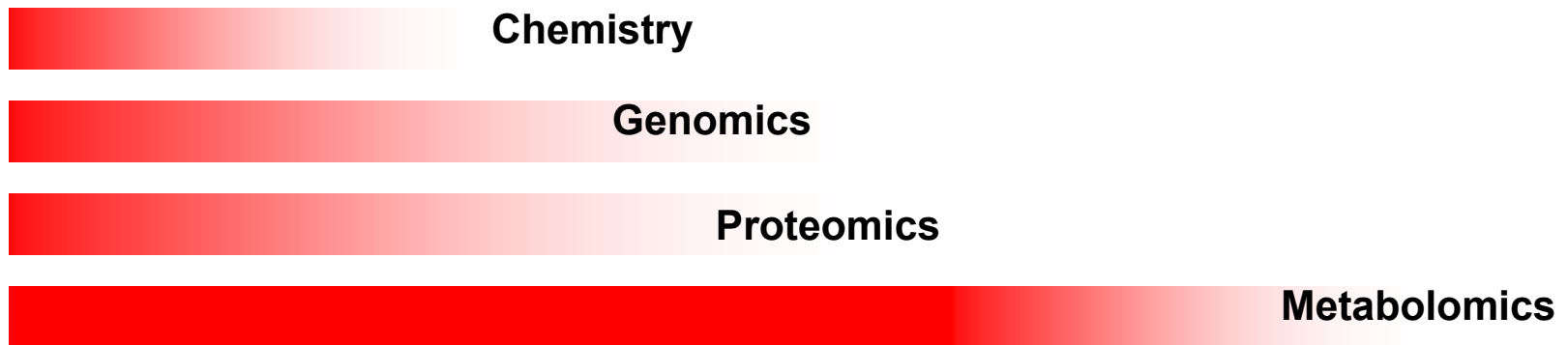


Key Trends in Metabolomics

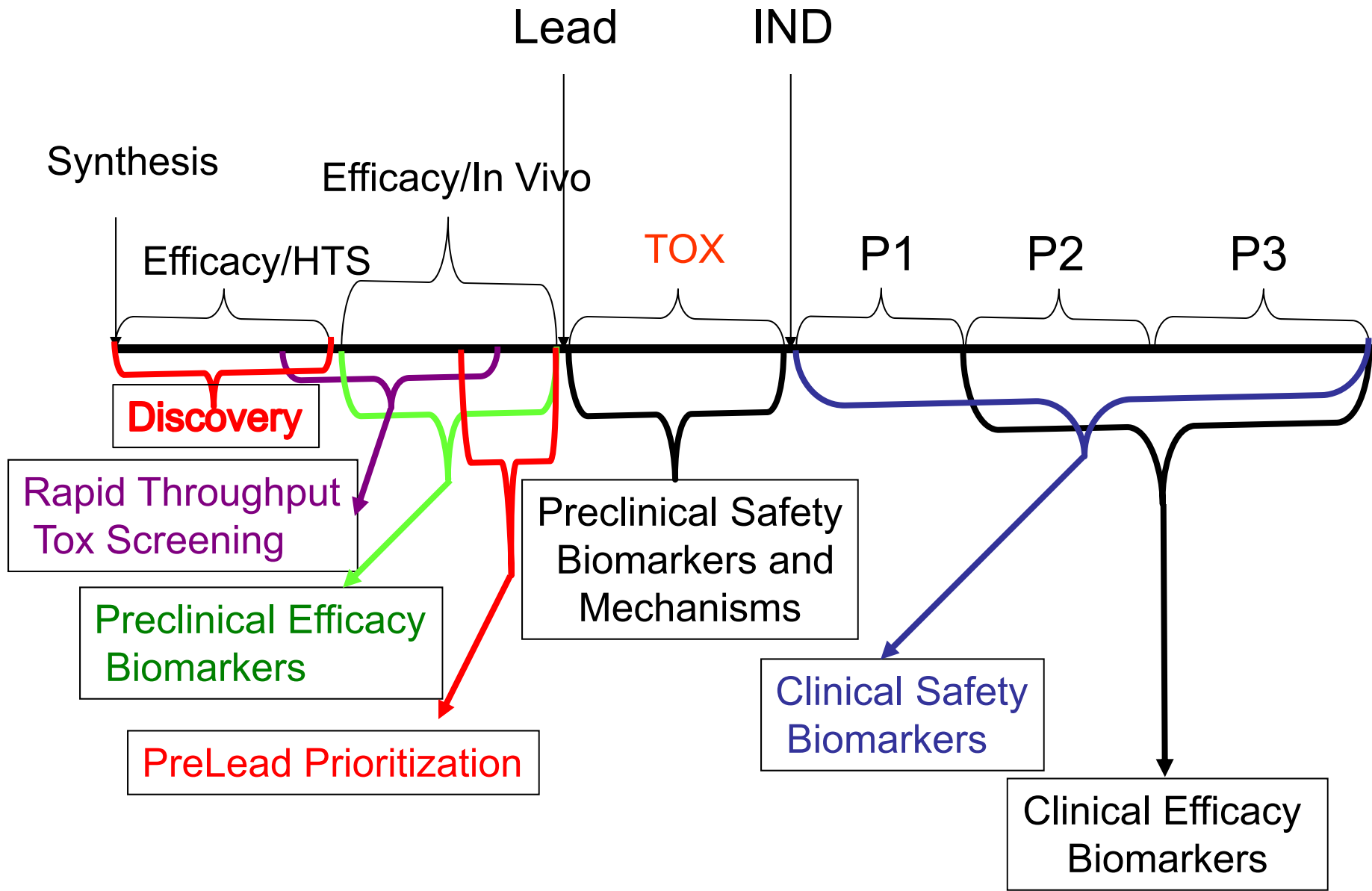
- Automated metabolomics
- Expanding metabolome coverage
- Making metabolomics portable
- Quantify, quantify, quantify...
- Moving metabolomics from the lab to the clinic
- Moving metabolomics (back) into drug development and discovery

Metabolomics & The Drug Industry

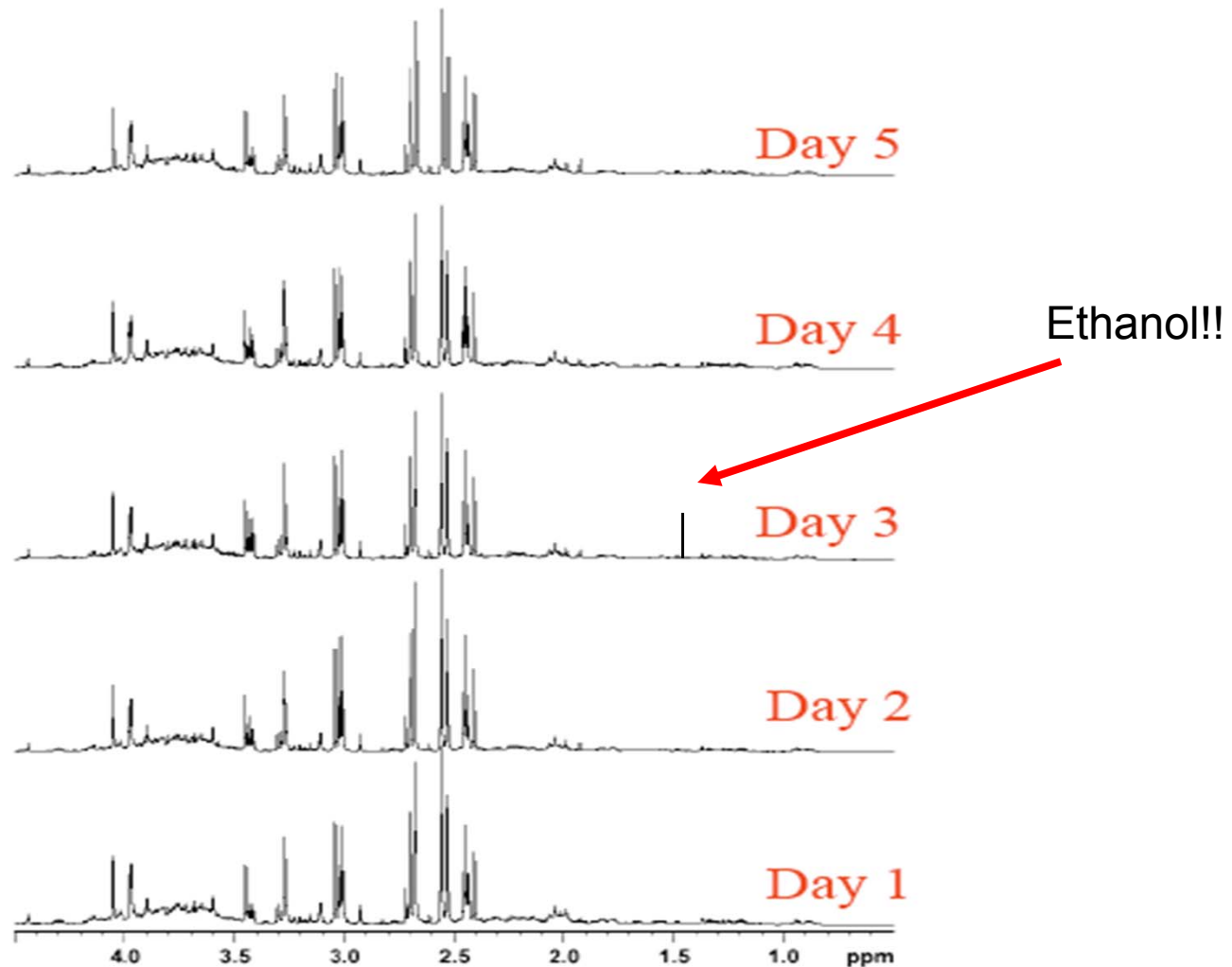
\$280	\$60	\$90	\$250	\$120 million
3.5 yrs	1 yr	2 yrs	3 yrs	2.5 yrs
Discovery	Phase I	Phase II	Phase III	FDA Approval



Metabolomics in Drug Development

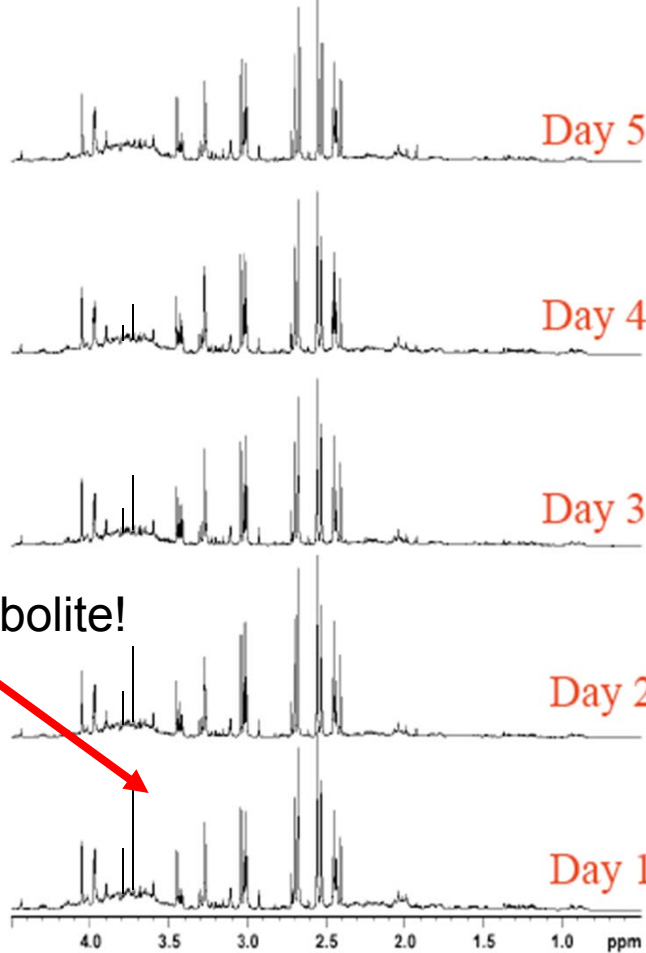


Applications in Patient Compliance (P2/P3 Trials)

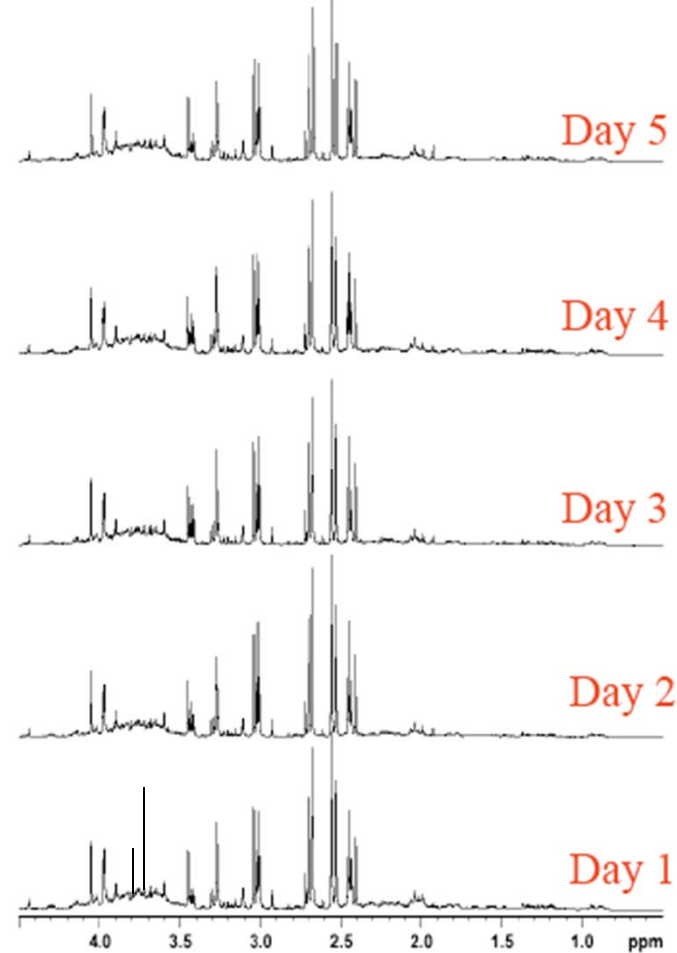


Applications in Drug Monitoring/Customization

Slow metabolizer

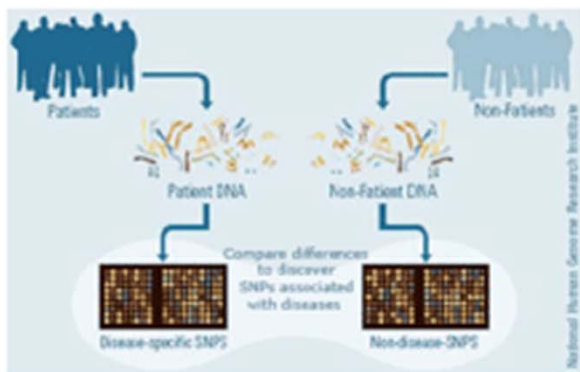


Fast metabolizer

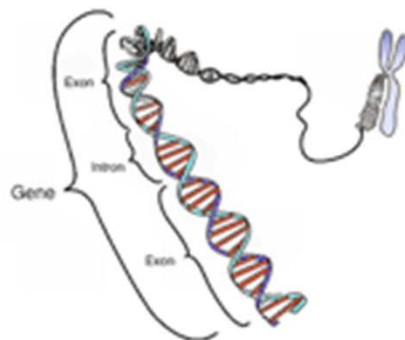


Traditional Drug Discovery

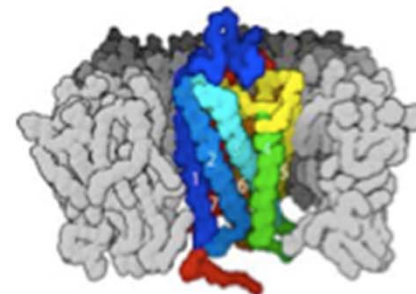
1:5 success \$2-10 million, 2-4 yrs



1:2 success



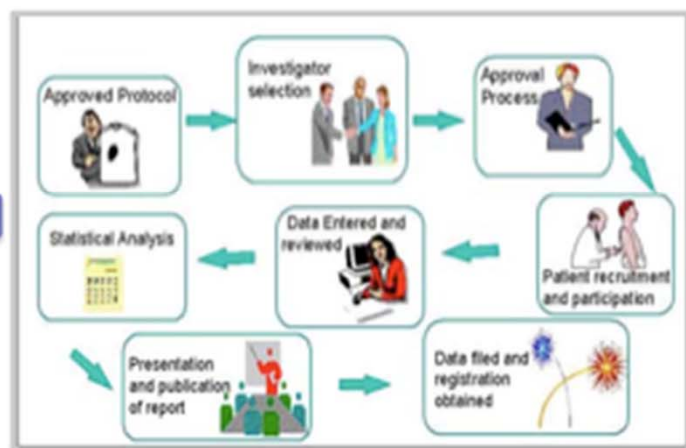
1:2 success



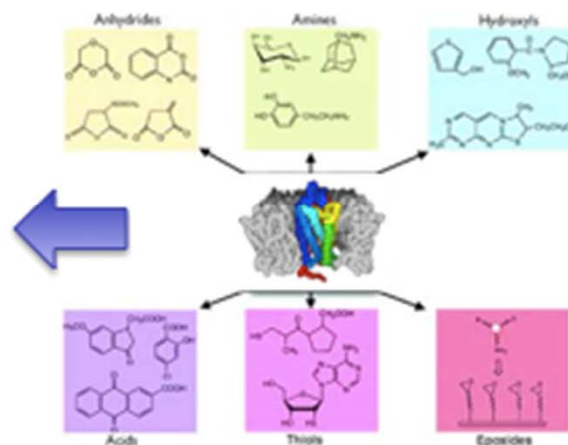
0.001% Success Rate, 20+ years, >\$1 billion



1:2 success



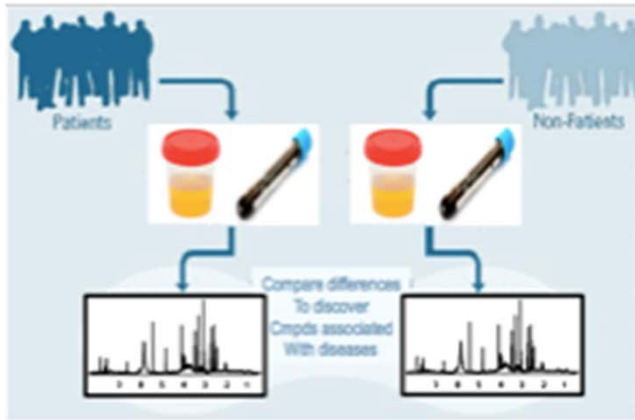
1:500 success, \$1 billion, 15 yrs



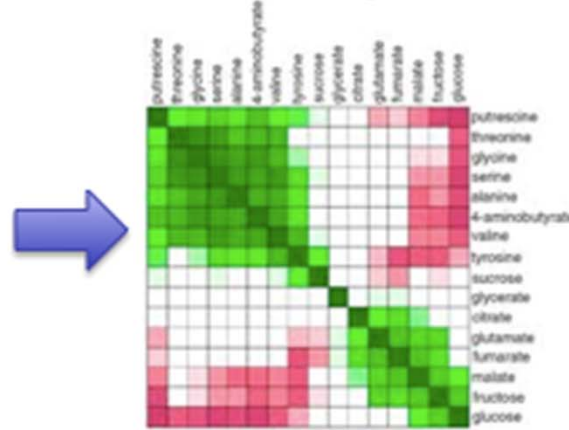
1:5 success, 1-5 yrs

Metabolite-Based Drug Discovery

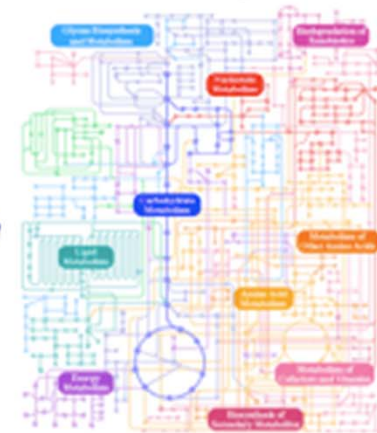
1:2 Success, \$200K, 1 yr



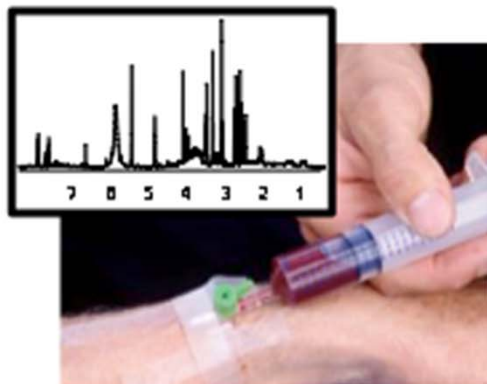
1-2 days



1-2 days



15% Success Rate, 1+ years, <\$250,000



1:2 Success, \$200/yr

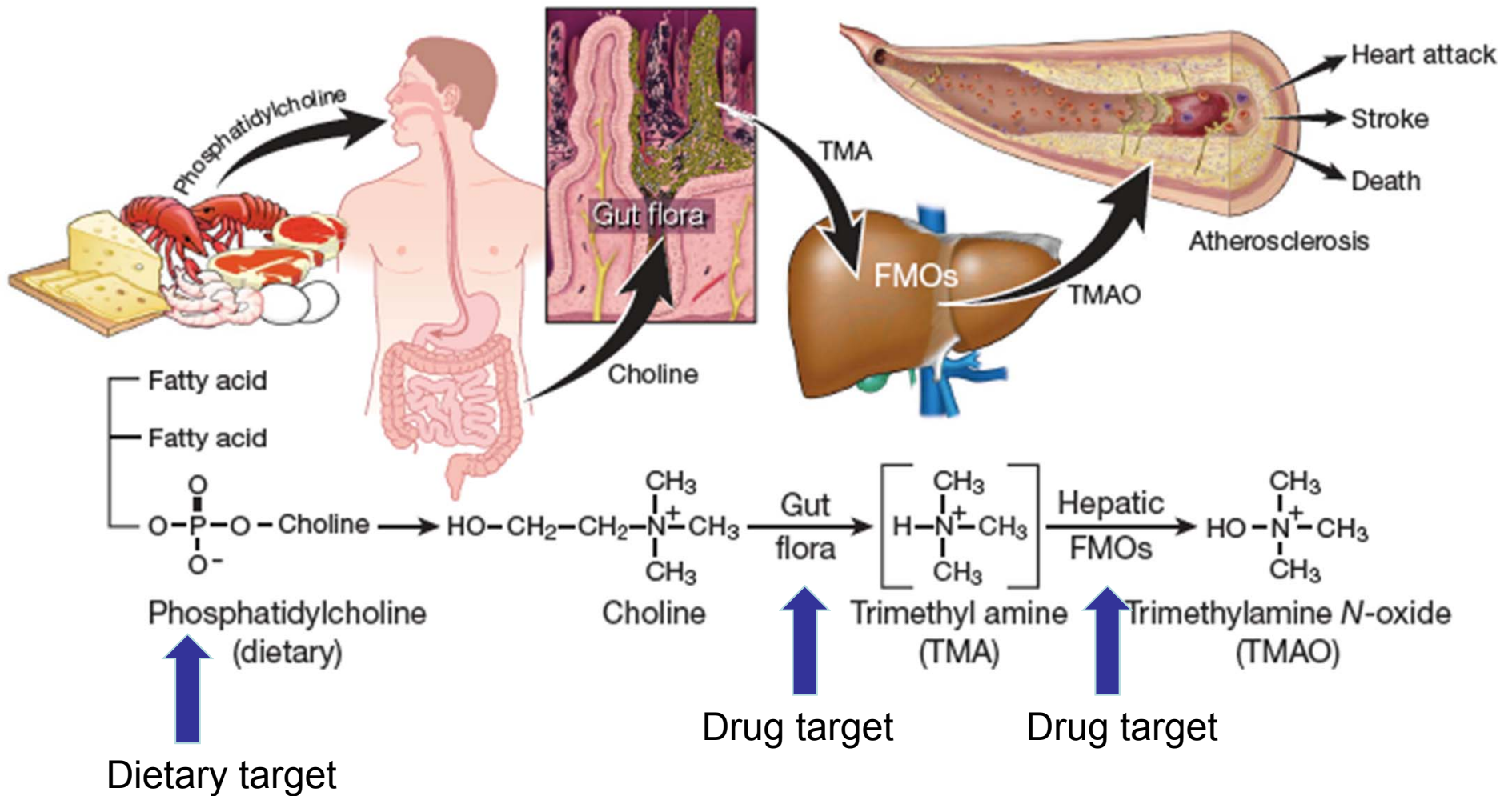


1-2 weeks (cmpds/diet)
5-10 yrs (enzymes, MAbs)



1:2 Success, 1-2 hours

Metabolomics, CVD & Therapy



Summary – The Future of Metabolomics

- **Automated metabolomics**
- **Expanding metabolome coverage**
- **Making metabolomics portable**
- **Quantify, quantify, quantify...**
- **Moving metabolomics from the lab to the clinic**
- **Moving metabolomics (back) into drug development and discovery**

Acknowledgements

- Augustin Scalbert
- Russ Greiner
- Jie Chen
- Khalid Aziat
- Felicity Allen
- Richard Fedorak
- Liang Li
- Rupa Mandal
- Jennifer Reid
- Tamara Lim
- Ray Bahado-Singh
- Jason Dyck
- David Rush
- Tom Blydt-Hansen
- Siamak Ravanbaksh
- Beomsoo Han
- Jeff Xia
- Tim Jewison
- Allison Pon
- Craig Knox



GenomeAlberta



GenomeCanada



CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada



**THE METABOLOMICS
INNOVATION CENTRE**