Mass Spec Data

As applied to metabolomics....

What is metabolomics?

- "Systemic study of the unique chemical fingerprints that specific cellular processes leave behind" *Bennett Daviss, The Scientist 19(8) April 2005*

- But…is not limited to cellular processes…
    toxins, drugs, endogenous versus non endogenous compounds

- …..and has great utility for……
    discovery research, translational science, clinical profiling, personalized medicine

Mass Spec and Metabolomics

The "Omics" Cascade

What can happen → **GENOME**

What appears to be happening → **TRANSCRIPTOME**

What makes it happen → **PROTEOME**

What has happened and is happening → **METABOLOME**

→ **PHENOTYPE**

1. **Sample collection**
   Treatment, disease, etc.   Control
   tissue, plasma, urine, saliva, cells, etc.

2. **Sample preparation**
   •Internal standard spike
   •Extraction (LLE, SPE, etc)

3. **Sample analysis**
   Derivatization-GC-MS, LC-MS, CE-MS
   Group 1   Group 2

4. **Data export**

Peak ID	Area(Height)				etc
	Sample 1a	Sample 1b	Sample 1c	Sample 2a	
0001	3500	3300	3000	7500	
0002	6000	5900	6100	3000	
0003	9000	8500	8600	5000	
0004	700	550	800	1500	
0005	1250	1100	1320	1310	
...	...	...	...	...	

5. **Data analysis**
   Multivariate analysis, e.g. PCA

   Score Plot (Groups)   Loading Plot (Markers)

6. **Marker Identification**
   Libraries, MS/MS, NMR

Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrom Rev. 2007 Jan-Feb;26(1):51-78. Review.

MS1 SEARCH ALGORITHM

Theoretical Features

C6H12O6?

C6H12O6

VS

C5H11O5

C2H15NO6P

...

ONE Experimental Feature

All Experimental Features from one sample

How different are the masses?

How different are the intensities?

Measured Difference

+

+

Single Good Hit

Single Bad Hit

Many Experimental Ions from a single sample compared to database

All Good and Bad Mass Hits

All Good and Bad Intensity Hits
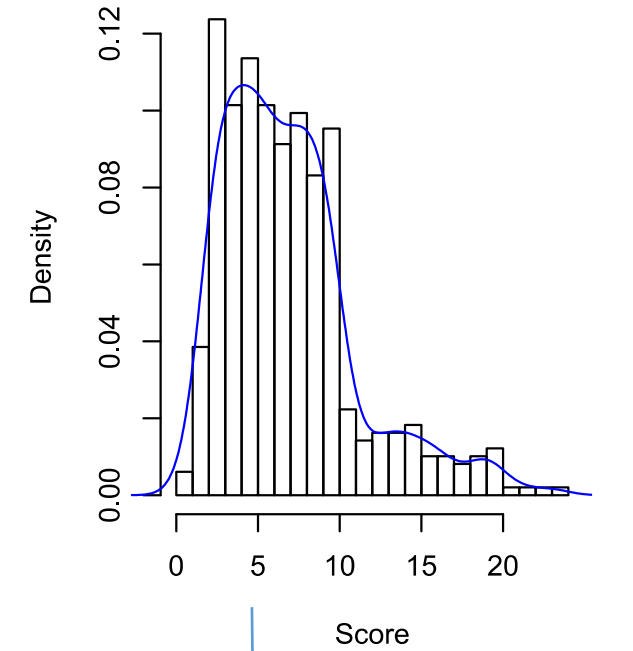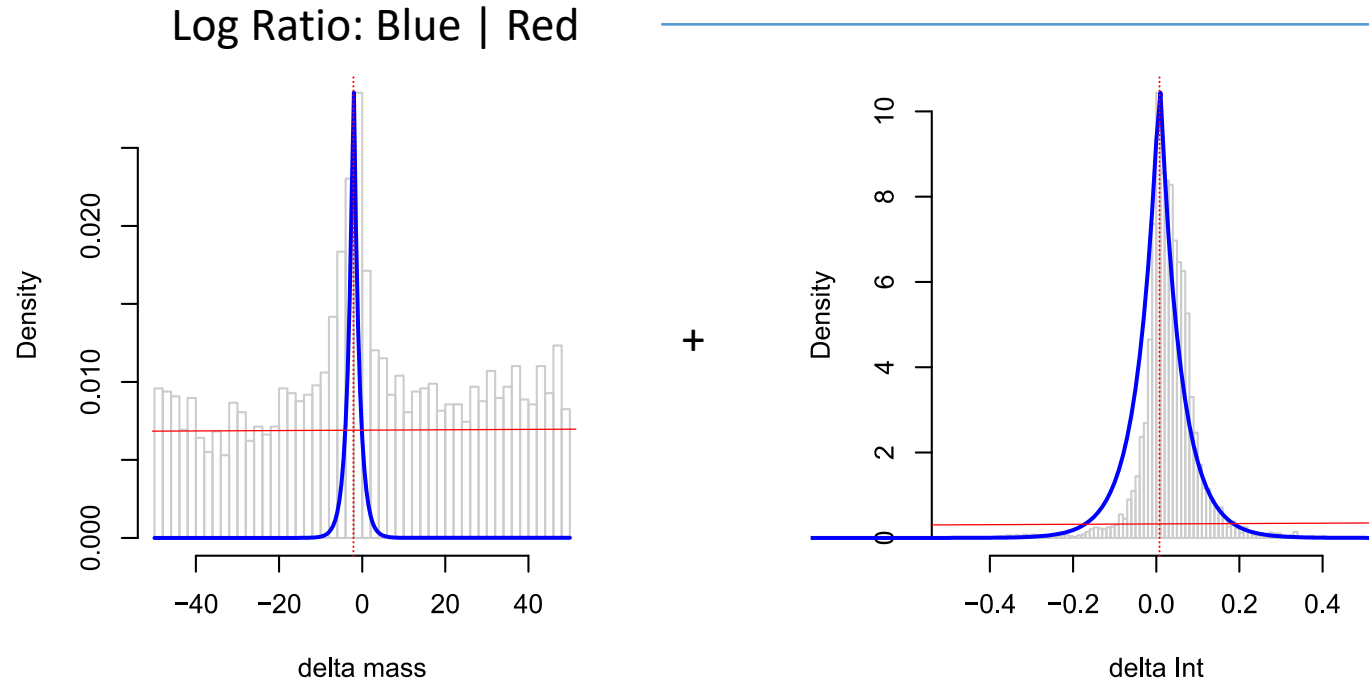
Difference in mass
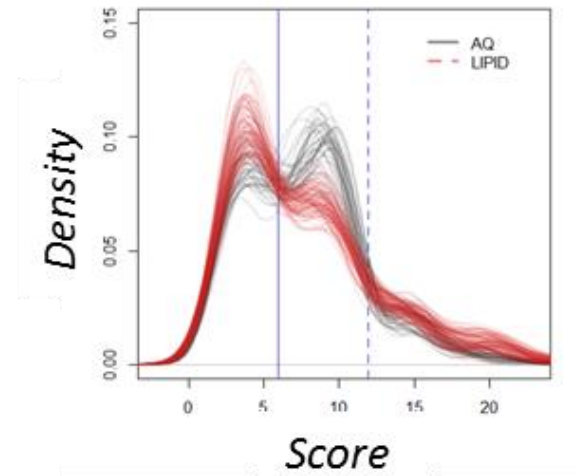
Difference in Intensity
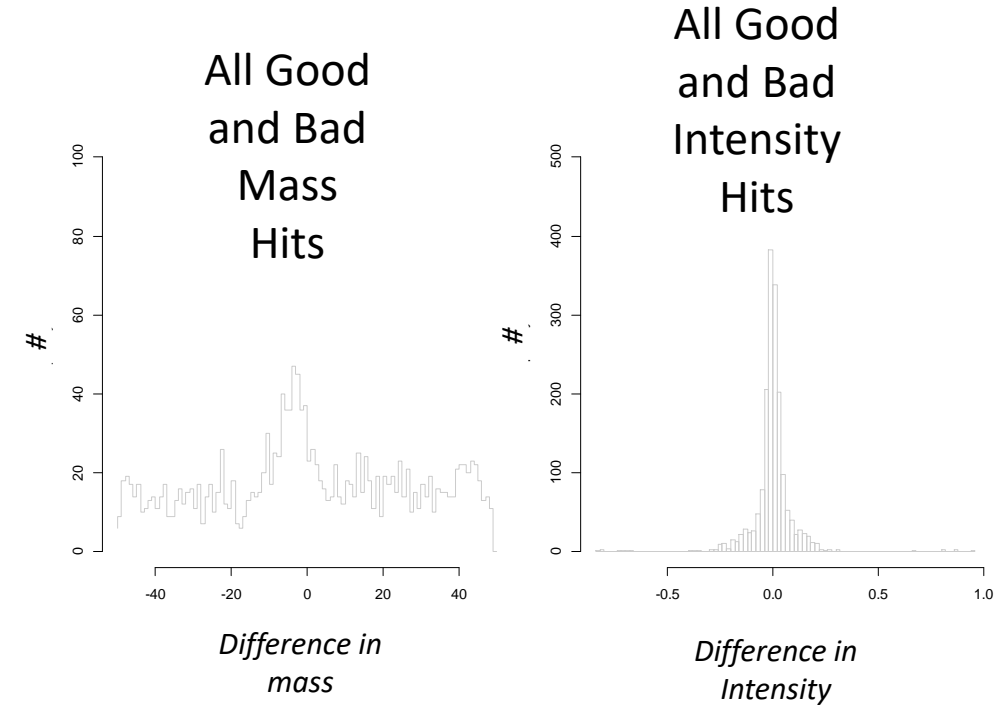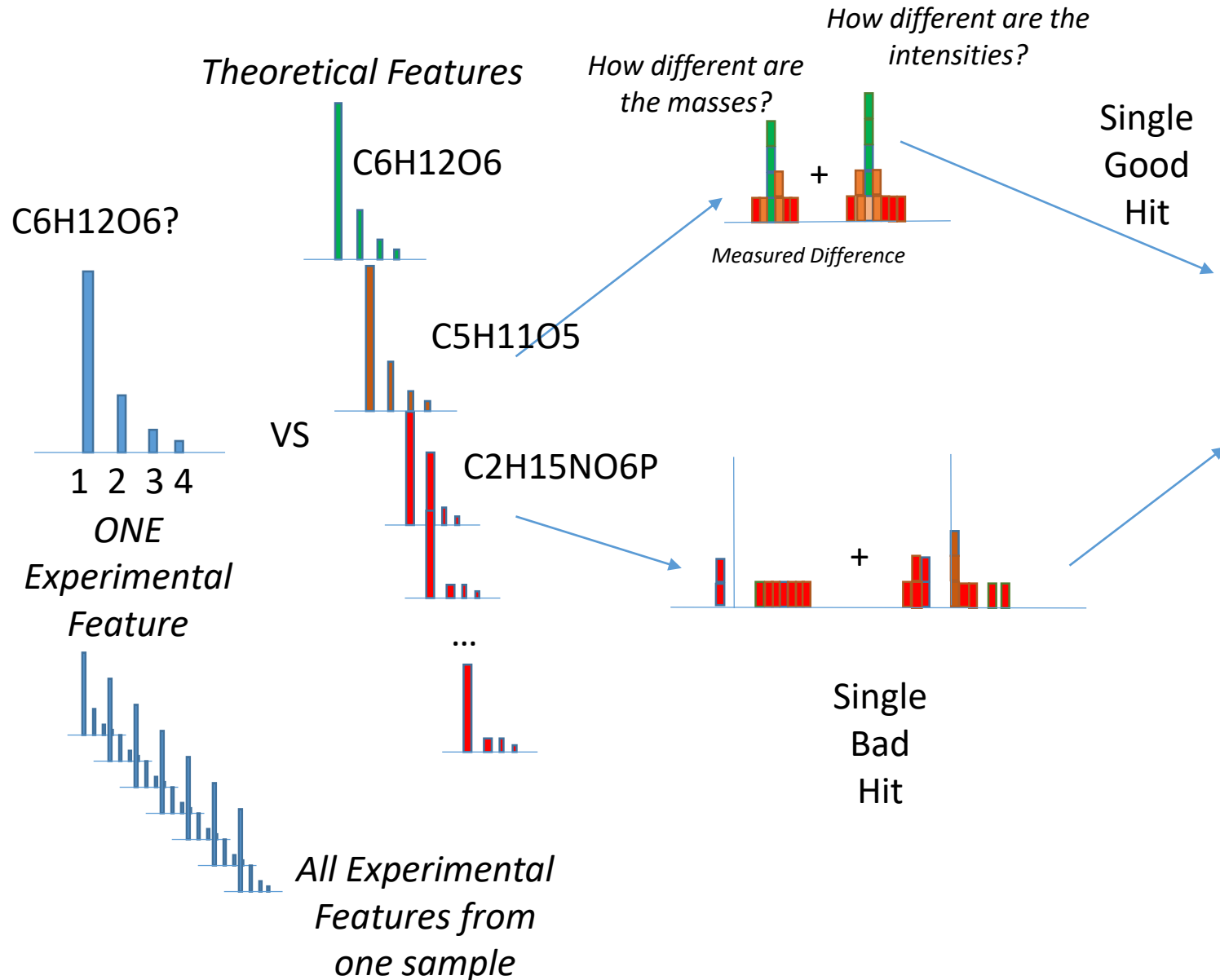
Fit curve to mass and intensity differences

MS1 SEARCH ALGORITHM

Curve fitting forms the premise for scoring.
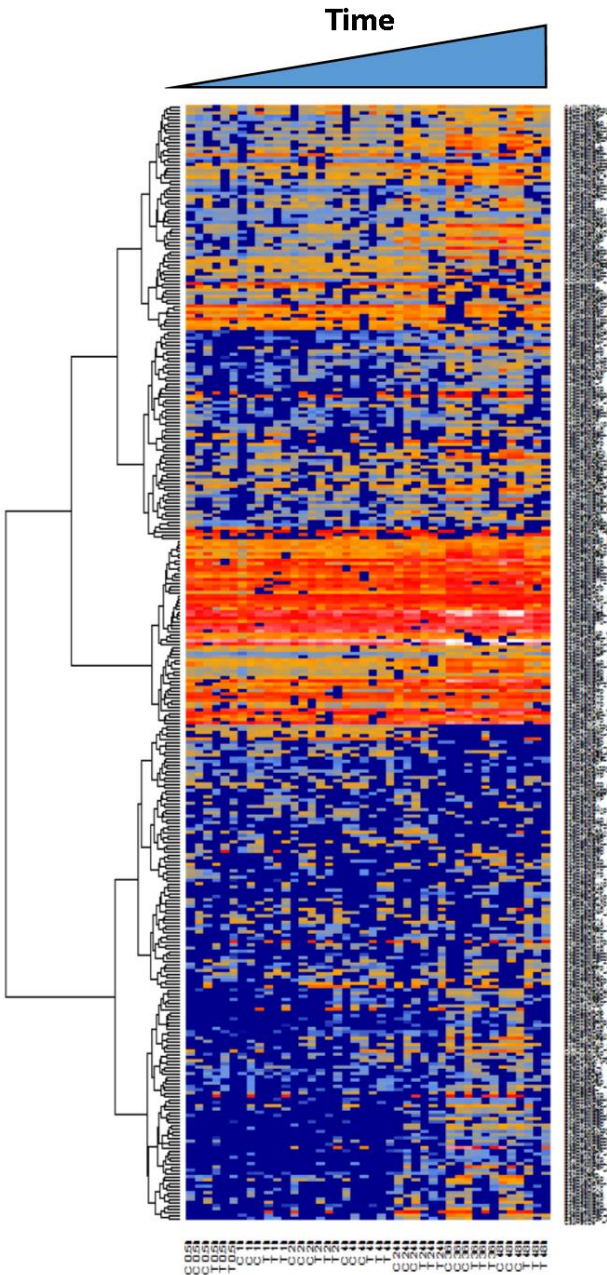*Score = Log ratio of delta masses + delta intensities*

Log Ratio: Blue | Red

Time

FDR 5 or 1%

Biological Response, Taxol Responsive
->Match expected biological outcome

Score and Intensity

Density Feature Finding 30day HEK293

Ceramide(d18:1/22:0)

Galabiosylceramide (d18:1 / 16:0)

Taxol

Control

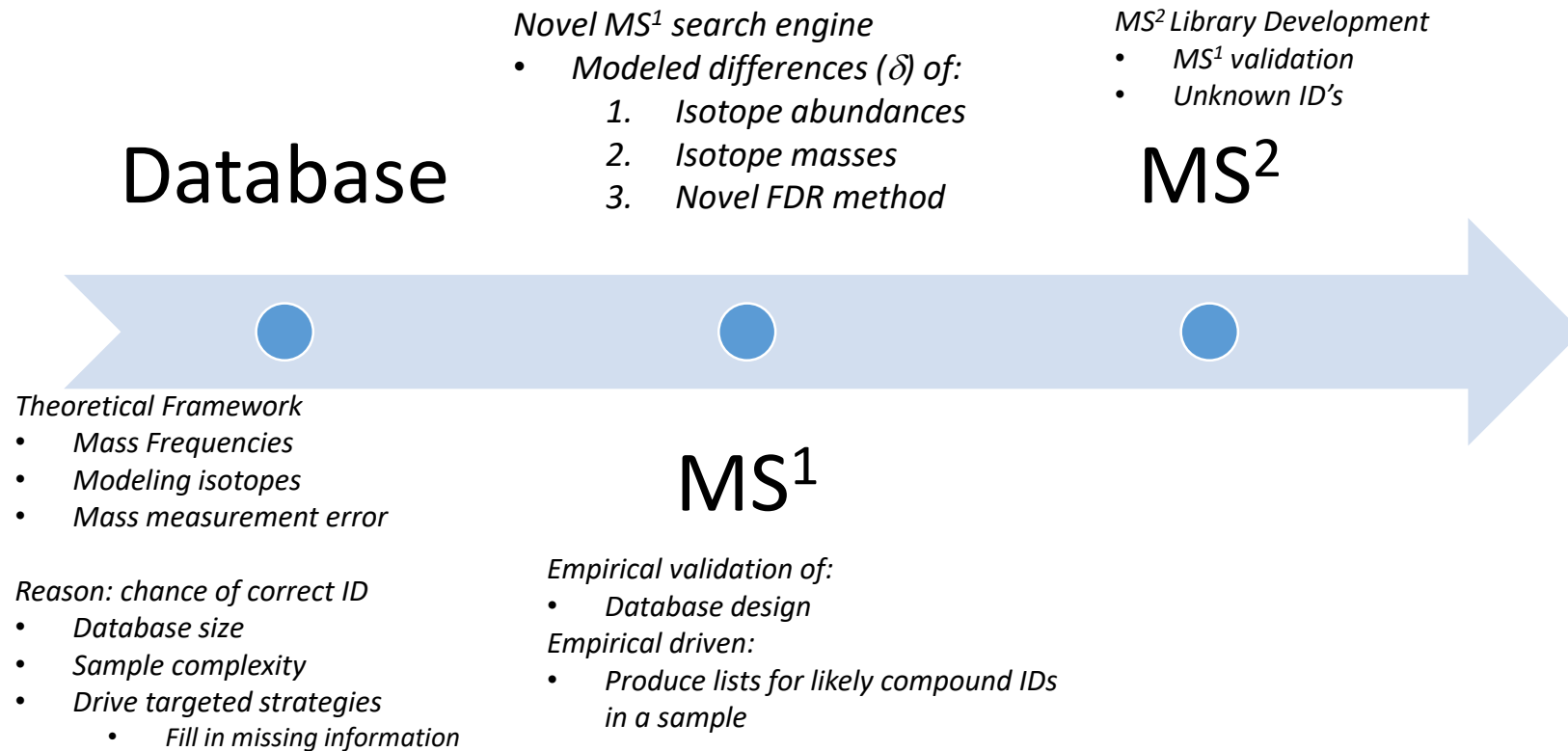Glucosylceramide

Sphingomyelin (d 18:0 / 16:1(9Z))

Sphingosine
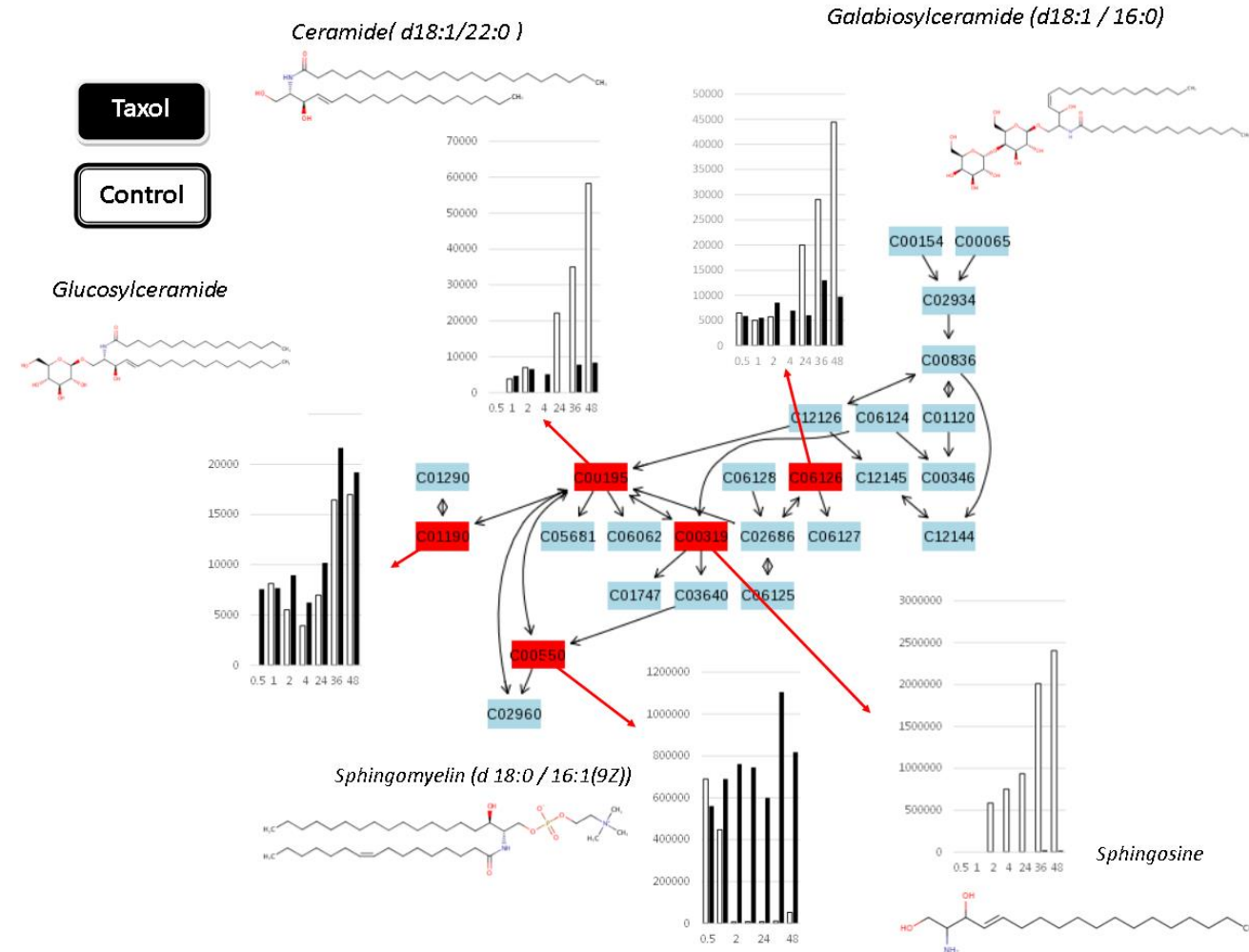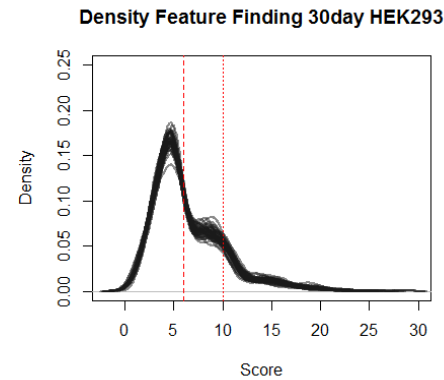
Sphingolipid Metabolism

Metabolomics: current vision

Challenges:
- Learning from proteomics
- Development of tools to enable rapid and accurate outcomes

Novel MS^1 search engine
- *Modeled differences (δ) of:*
  1. *Isotope abundances*
  2. *Isotope masses*
  3. *Novel FDR method*

MS^2 Library Development
- *MS^1 validation*
- *Unknown ID's*

Database

MS^2

MS^1

Theoretical Framework
- *Mass Frequencies*
- *Modeling isotopes*
- *Mass measurement error*

Reason: chance of correct ID
- *Database size*
- *Sample complexity*
- *Drive targeted strategies*
  - *Fill in missing information*

Empirical validation of:
- *Database design*

Empirical driven:
- *Produce lists for likely compound IDs in a sample*

Proteomics vs metabolomics key publications and dates

Publications by Year
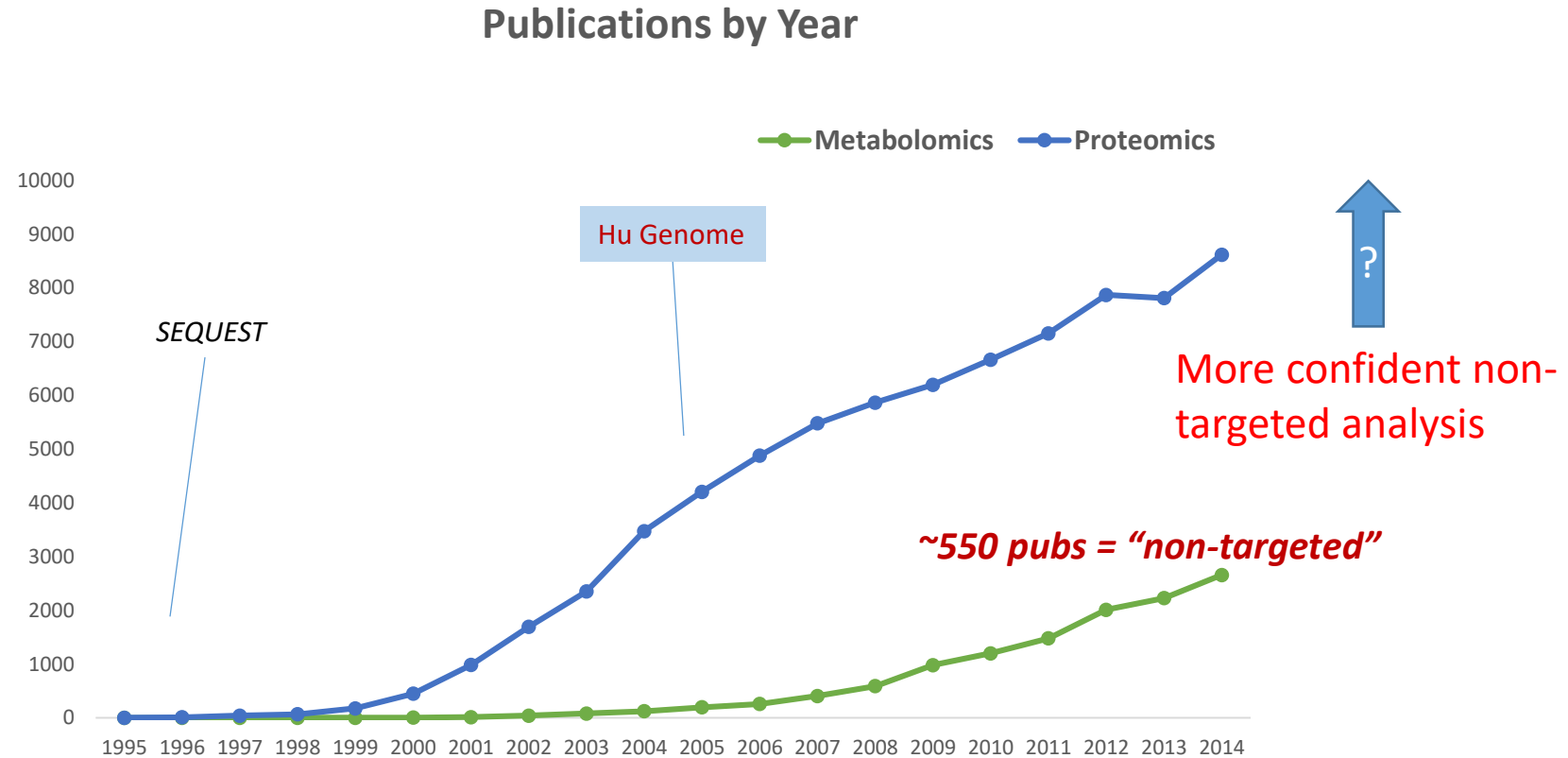
Identifying challenges appropriate for the effort in metabolomics

Sweet Fruit — Intelligent Methods for Quantitation

Bulk of Fruit — *MS¹ Search Engine Probabilistic Models

Low-Hanging Fruit — Re-design of Proteomics tools *Statistical Tools for LCMS QA *Spectral Libraries

Ground Fruit — Re-invention of already useful tools Past trends
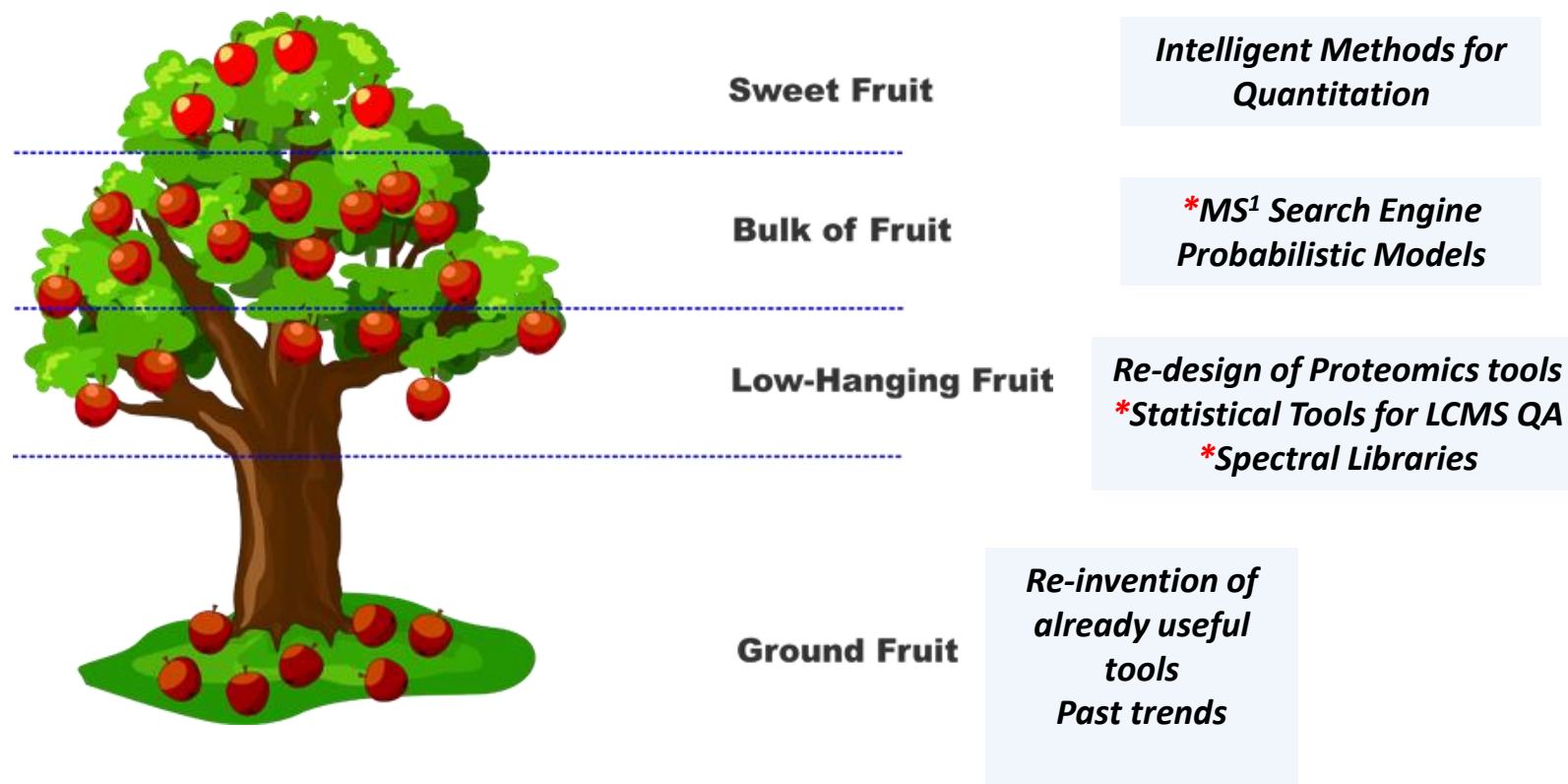
Future Plans

- Produce broadly impacting set of tools available to the discipline
- Scoring algorithms for mass spec
- Pathways: inferring greatest success of identity of a pathway
- User friendly interfaces for data viewing

Collaborate

MATH

BIOCHEM

COMP.
Science