Icahn School of Medicine at
Mount Sinai

Reactomics: Using mass spectrometry as a chemical reaction detector

Miao Yu
2018/11/11 (updated: 2020-01-30)

Workflow for metabolomics

- Collect samples
- Collect peaks from mass spectrometry
- Annotate peaks to get the compound name
- Build the links between compounds by pathway/network analysis

> Sample -> Peaks -> Compounds -> Relationship among compounds

- Problems
  - Time consuming - too many peaks ~20k
  - Sensitivity - low for full scan
  - Standards coverage - unknown unknown

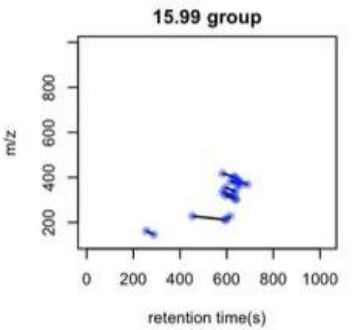
My Idea

> Sample -> Peaks -> ~~Compounds~~ -> Relationship among compounds

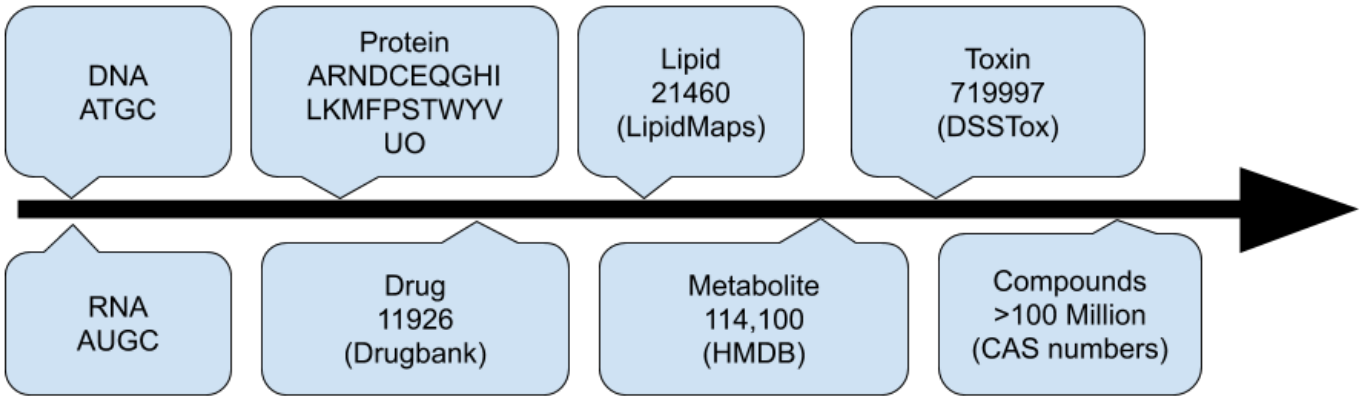
- Mass spectrum could directly measure relationsship(reactions)

$$[M_{1,2,...,n}+H]^+ \xleftrightarrow{\Delta 16\ amu} [M_{1,2,...,n}+OH]^+$$

- Xenobiotics Phase I Reactions
  - Hydrolysis
  - Reduction
  - Oxidation



Why Reactions?



- Unit: Gene(5) < Protein(20+2) < Metabolite(100K) < Compound(100M)
- Combination: Gene(20,000-25,000) < Protein(20,000-25,000) < Compound(???)
- Small molecular **combination** is chemical reaction or paired mass distance

Why PMD?

- Nuclear Binding Energy

$$\Delta m = Zm_H + Nm_n - M$$

- The missing mass was converted into energy ($E = mc^2$) and emitted when the atom made
- Atoms -> Compounds -> Mass distances between compounds
- **Paired Mass Distances(PMD)** is unique
- **High resolution** mass spectrometry WINs

Sources of PMDs in the real data

Where is PMD?

- Isotopologues
  - $[M]^+$ $[M+1]^+$
  - 1.006 Da
- in source reaction
  - $[M+H]^+$ $[M+Na]^+$
  - 21.982 Da
- Homologous series
  - Lipid $-[CH_2]-$
  - 14.016 Da
- Xenobiotic metabolism
  - Phase I hydrolation
  - 15.995 Da

Quantitative and Qualitative analysis for Reaction

KEGG reaction database

PMD	Freq	Example
1.008	2037	NAD(+) + succinate <=> fumarate + H(+) + NADH
2.016	1748	NAD(+) + propanoyl-CoA <=> acryloyl-CoA + H(+) + NADH
15.995	1170	ATP + GDP <=> ADP + GTP
13.979	1122	deoxynogalonate + O2 <=> H(+) + H2O + nogalonate
17.003	929	H2O + hypotaurine + NAD(+) <=> H(+) + NADH + taurine
79.966	750	ATP + H2O <=> ADP + H(+) + phosphate
14.016	611	acetyl-CoA + propanoate <=> acetate + propanoyl-CoA
0	533	L-glutamate <=> D-glutamate
162.053	365	H2O + lactose <=> D-galactose + D-glucose
18.011	361	L-serine <=> 2-aminoprop-2-enoate + H2O

- Real reactions contain ions
- Biased by known reactions

Quantitative and Qualitative analysis for Reaction

HMDB compounds database

	C	H	O
14.016	1	2	0
2.016	0	2	0
28.031	2	4	0
26.016	2	2	0
15.995	0	0	1
12	1	0	0
56.063	4	8	0
42.047	3	6	0
30.011	1	2	1
24	2	0	0

- Dominated by C, H and O
- Structure or reaction?

Quantitative and Qualitative analysis for Reaction

HMDB compounds database

	PMD	frequency	accuracy	PMD	frequency	accuracy
+C2H	14.016	4934	0.9755	14.02	8003	0.6014
+2H	2.016	4909	0.9703	2.02	7959	0.5984
+2C4H	28.031	4878	0.9783	28.03	7799	0.6119
+2C2H	26.016	4229	0.9775	26.02	7343	0.5630
+O	15.995	4214	0.9808	15.99	7731	0.5346
+C	12.000	3861	0.9826	12.00	7145	0.5310
+4C8H	56.063	3861	0.9653	56.06	6699	0.5564
+3C6H	42.047	3771	0.9737	42.05	6558	0.5599
+C2HO	30.011	3698	0.9440	30.01	6761	0.5163
+2C	24.000	3689	0.9810	24.00	6963	0.5197

Quantitative and Qualitative analysis for Reaction

HMDB compounds database

	PMD	frequency	accuracy	PMD	frequency	accuracy
+C2H	14.0	50419	0.0955	14	156245	0.0354
+2H	2.0	50467	0.0944	2	156260	0.0352
+2C4H	28.0	50797	0.0939	28	155410	0.0356
+2C2H	26.0	48517	0.0852	26	154346	0.0309
+O	16.0	51278	0.0806	16	155811	0.0307
+C	12.0	49335	0.0769	12	155339	0.0283
+4C8H	56.1	36417	0.1026	56	151894	0.0286
+3C6H	42.0	49808	0.0737	42	153764	0.0275
+C2HO	30.0	51241	0.0681	30	154369	0.0260
+2C	24.0	48099	0.0752	24	154278	0.0273

Quantitative and Qualitative analysis for Reaction

Static v.s. dynamic

- Static mass pairs: paired intensity ratio is stable across samples
- Dynamic mass pairs: paired intensity ratio is stable across samples
- For example, [A,B], [C,D] and [E,F] are involved in the same PMD:

A	B	Ins ratio	C	D	Ins ratio	E	F	Ins ratio
100	50	2:1	100	50	2:1	30	40	3:4
1000	500	2:1	10	95	2:19	120	160	3:4

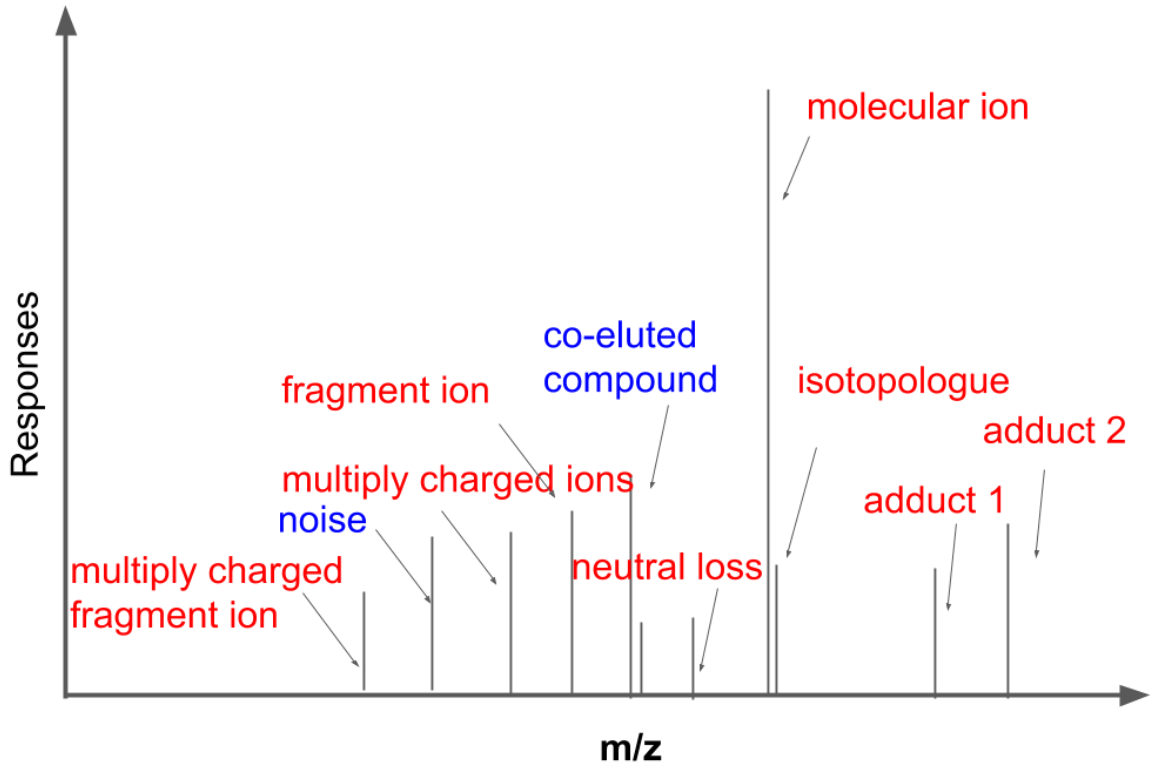
- [A,B] and [E,F] could be used for Quantitative analysis for certain PMD, rsd cutoff 30%
- [C,D] could be used to check dynamics of specific reaction

Reactomics Application

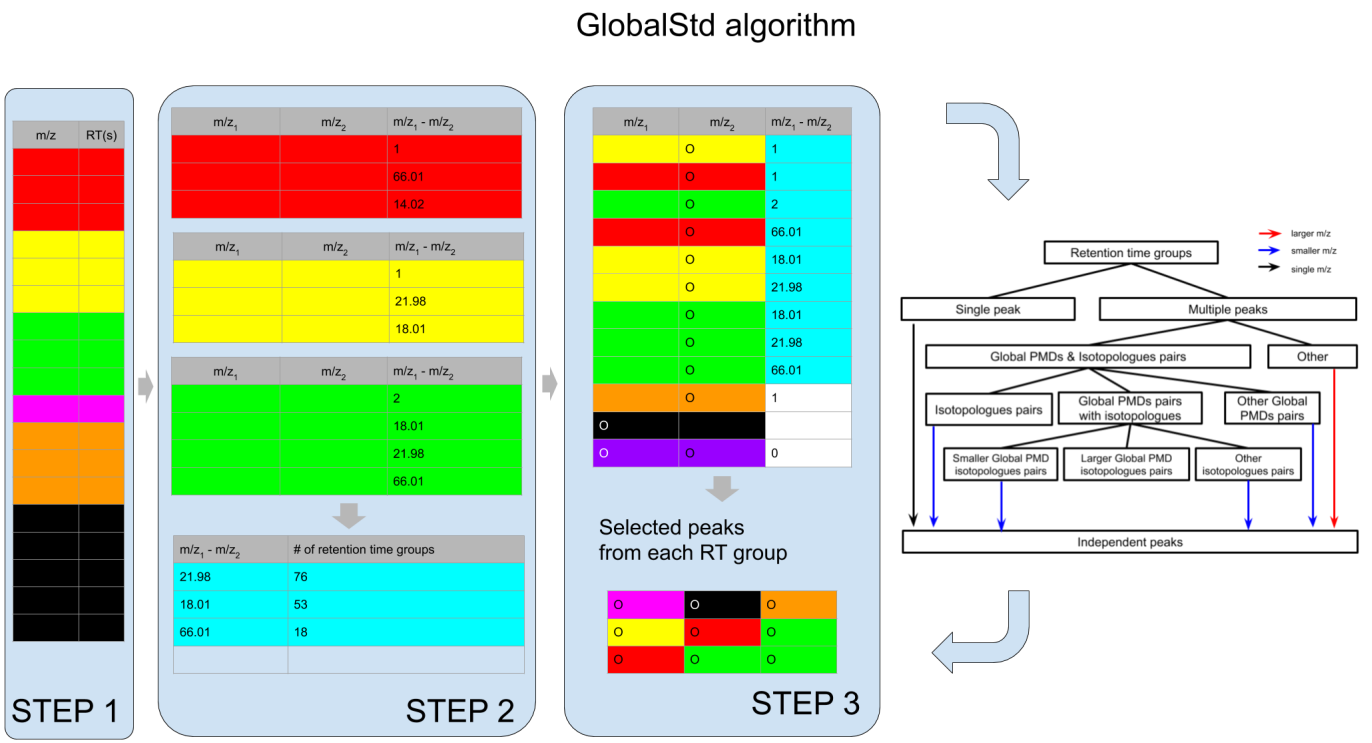
Global screen

Redundant peaks

- Untargeted analysis would loss sensitivity to capture all peaks
- Send unknown while independent peaks for MS/MS



GlobalStd Algorithm



Yu, M., Olkowicz, M., & Pawliszyn, J. (2019). Structure/reaction directed analysis for LC-MS based untargeted analysis. Analytica Chimica Acta, 1050, 16–24. doi:10.1016/j.aca.2018.10.062

Target compounds validation

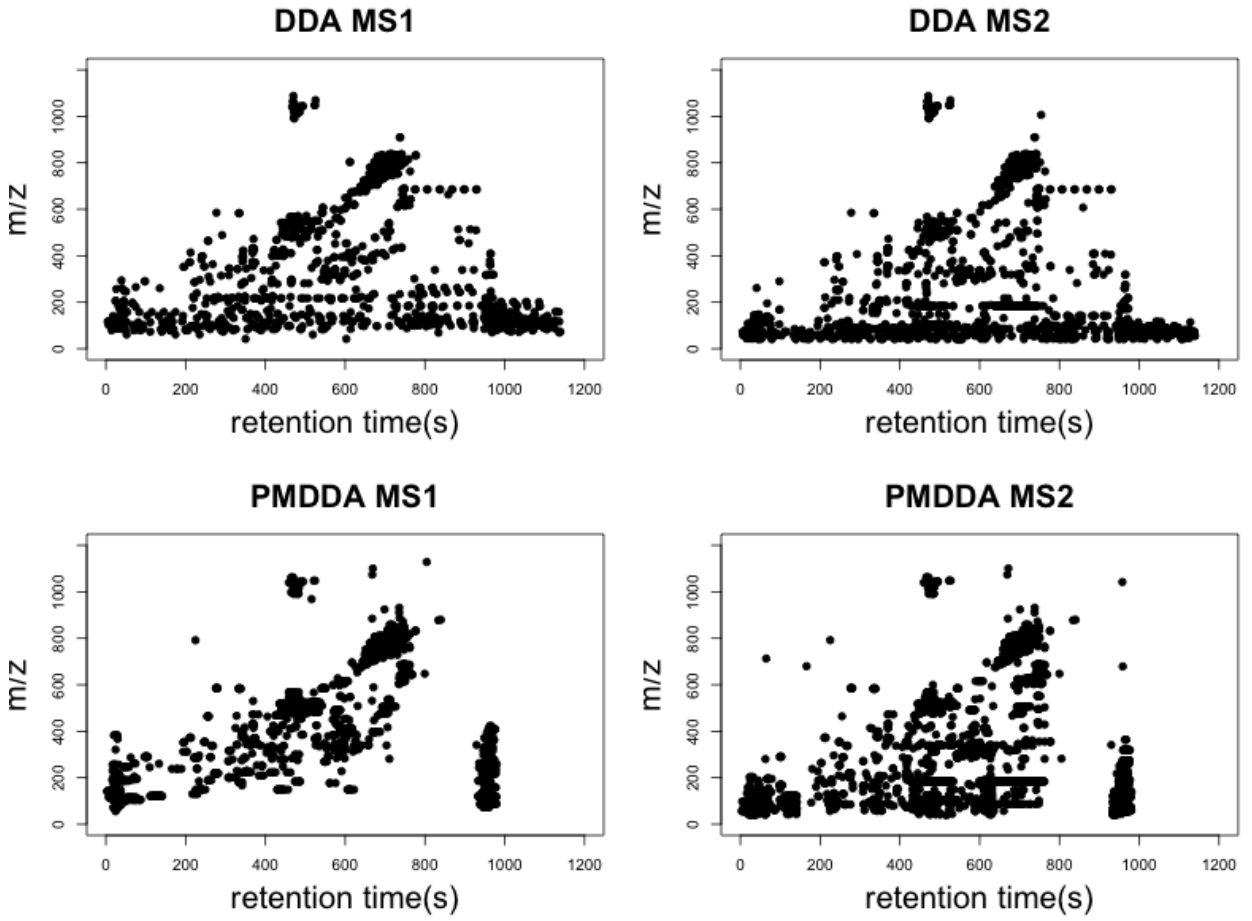
	Independent peaks	Target compounds found
pmd	985	18
CAMERA	1297	15
RAMclust	461	12
profinder	6628	7

- 103 compounds for validation
- 36 compounds could be found by xcms 6885 features
- 7 could be found by profinder untargeted analysis 6628 features

Untargeted MS/MS analysis - PMDDA

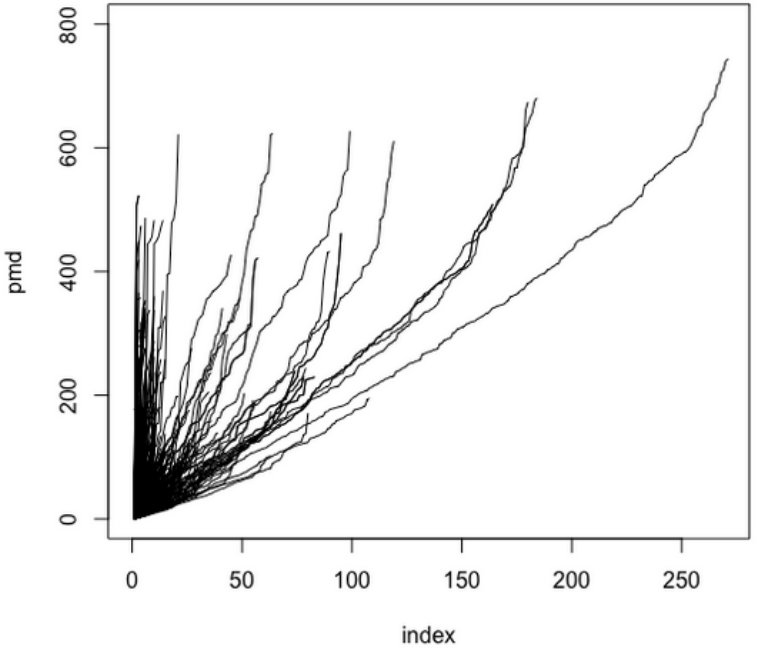
- Only use GlobalStd peaks for MS/MS analysis
  - Multiple injections
- MS/MS spectral library annotation on GNPS
- Compare with Data Dependent Acquisition (DDA) (173 compounds)
  - Annotated 235 extra compounds and overlap 59 compounds
  - Less contaminant ions

Untargeted MS/MS analysis - PMDDA



Untargeted MS/MS analysis - PMD Annotation

- Use pmd and rank of pmd for annotation
- Intensity filter(10%) and robust for noise
- 957/1098(87%) PMDR/HMDB QqQ data
- some compounds share the same pmd



Reactomics Application

Metabolites Discovery

Metabolites of exogenous compound

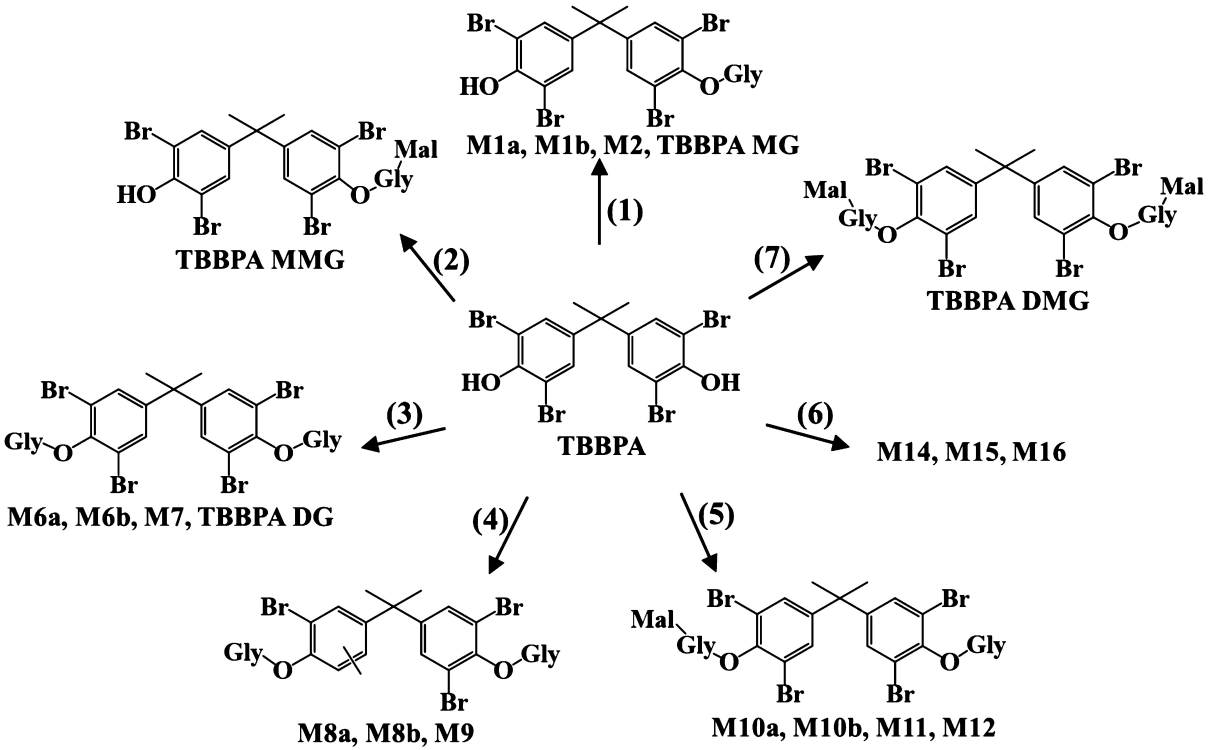
- Environmental pollution metabolites
- Drug metabolites

Xenobiotic metabolism

- Phase I
  - Oxidation (R-H ⇒ R-OH, pmd 15.995 Da)
  - Reduction (R-C=O ⇒ R-C-OH, pmd 2.016 Da)
- Phase II
  - Methylation (R-OH ⇒ R-O-C,pmd 14.016 Da)
  - Sulfation (R-OH ⇒ R-SO4, pmd 46.976 Da)
  - Acetylation (R-OH ⇒ R-O-COCH3, pmd 42.011 Da)
  - Glucuronidation (R-NH2 ⇒ R-NH-C6H9O7, pmd 192.027 Da)
  - Glycosylation (R-OH ⇒ R-O-C6H11O5, pmd 162.053 Da)

Metabolites of TBBPA in Pumpkin

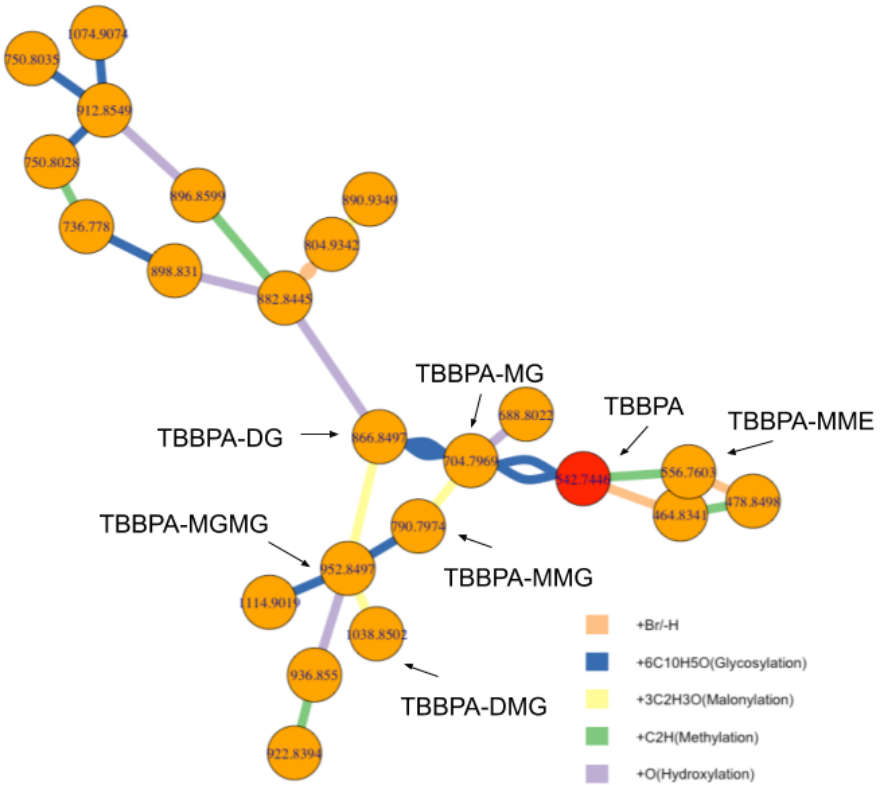
- Mass defect analysis to screen Brominated Compounds
- Confirmation by synthesized standards



Hou, X., Yu, M., Liu, A., Wang, X., Li, Y., Liu, J., … Jiang, G. (2019). Glycosylation of Tetrabromobisphenol A in Pumpkin. Environmental Science & Technology, 53(15), 8805–8812. doi:10.1021/acs.est.9b02122

Metabolites of TBBPA in Pumpkin

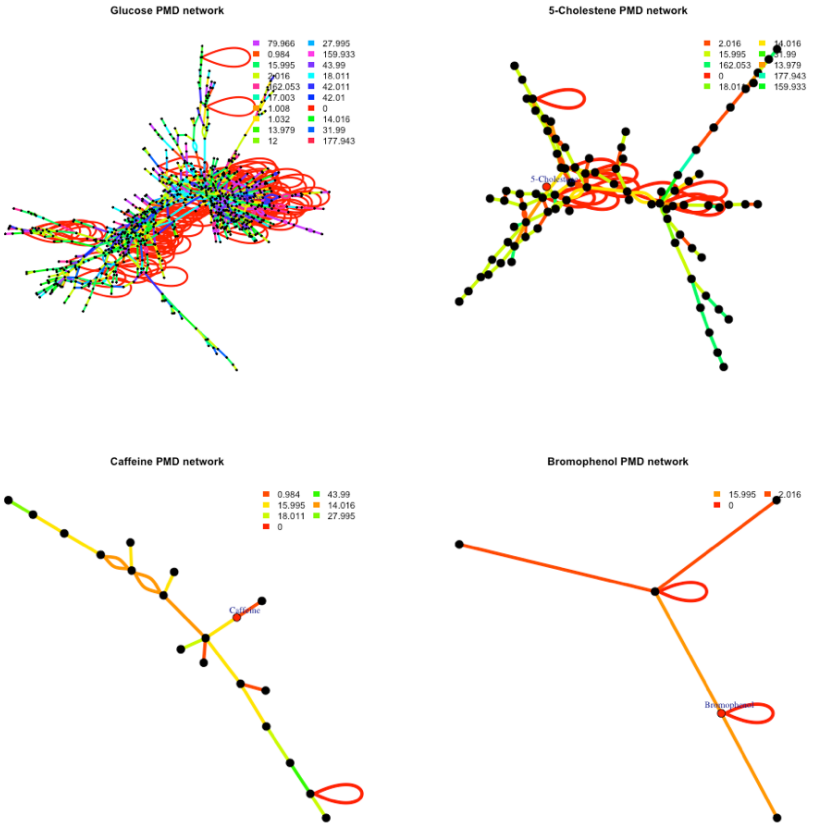
- TBBPA Metabolites PMD network



Hou, X., Yu, M., Liu, A., Wang, X., Li, Y., Liu, J., … Jiang, G. (2019). Glycosylation of Tetrabromobisphenol A in Pumpkin. Environmental Science & Technology, 53(15), 8805–8812. doi:10.1021/acs.est.9b02122

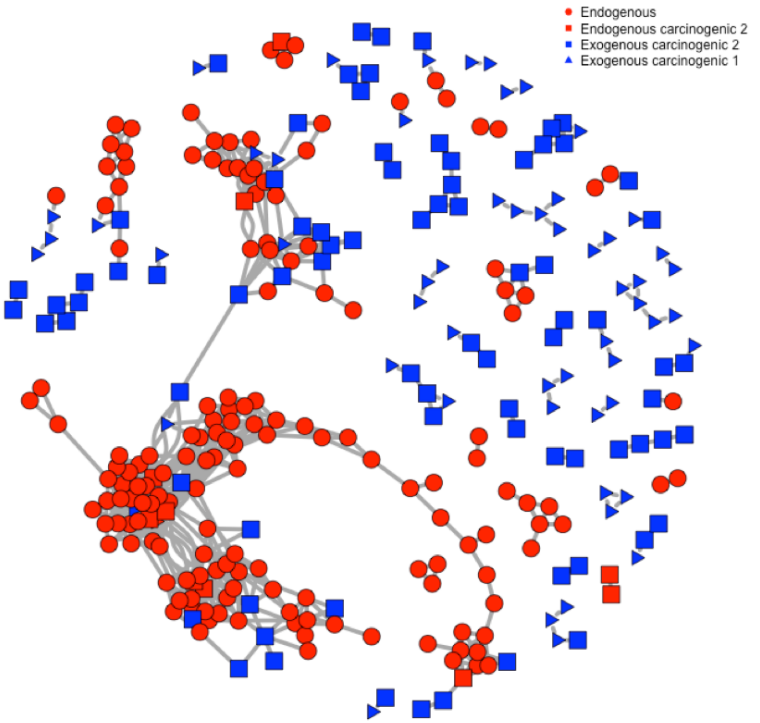
KEGG reaction network

- Metabolites of four compounds



Endogenous vs Exogenous

- T3DB Endogenous (255) vs Exogenous (705)
- Use top 20 high frequency PMDs

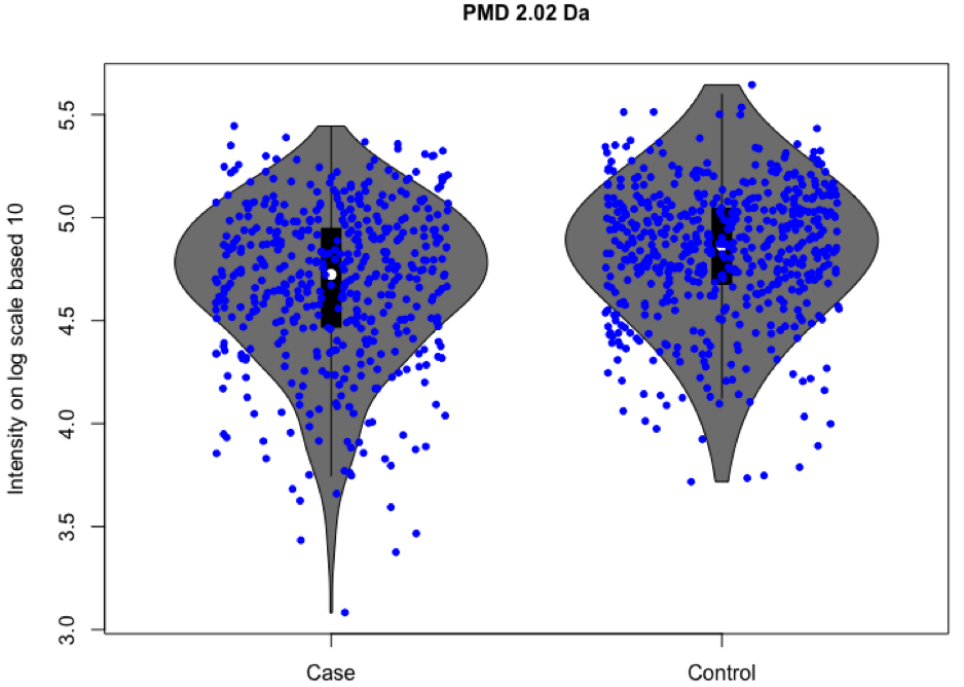


Reactomics Application

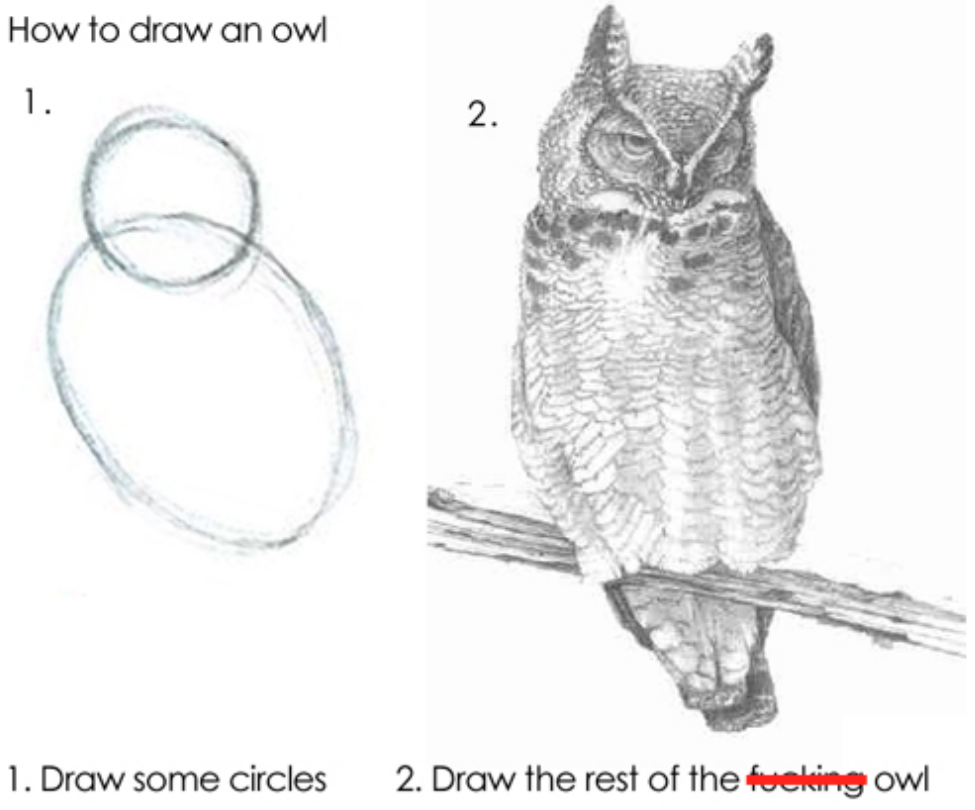
Biomarker Reaction

Lung cancer

- MTBLS28 1005 human urine samples
- PMD 2.02 Da show differences among control and diseases



How



Paper method v.s. Practical method in Metabolomics

Software

enviGCMS package

- Target analysis
- Mass defect analysis

pmd package

- Untargeted analysis
- GlobalStd algorithm
- Reactomics analysis

rmwf package

- NIST 1950 data
- Script

Thanks

Q&A

miao.yu@mssm.edu