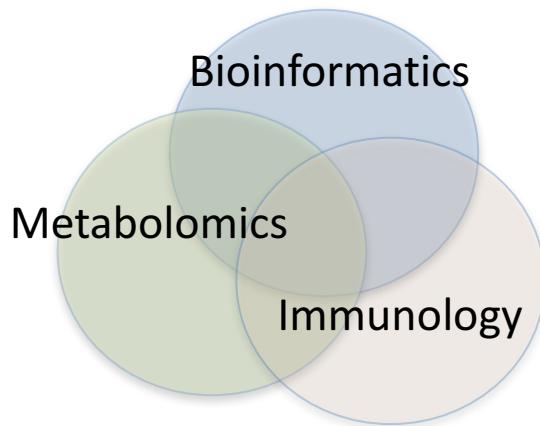




Overview of Metabolomics Workflow



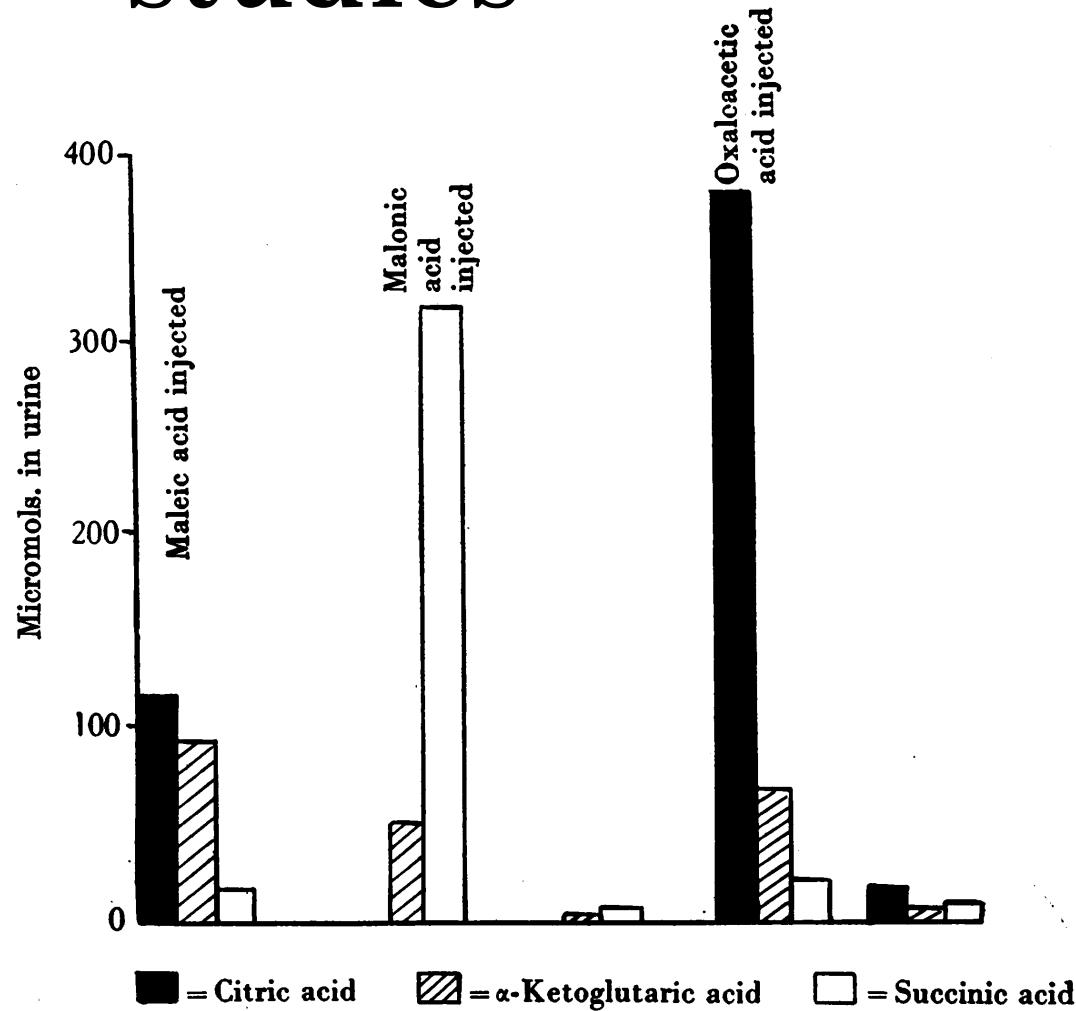
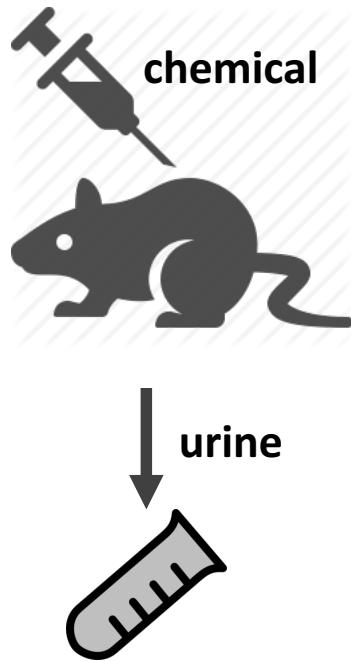
Shuzhao Li, Ph.D
Assistant Professor
Department of Medicine
Emory University
August 26, 2019

Full course material is at

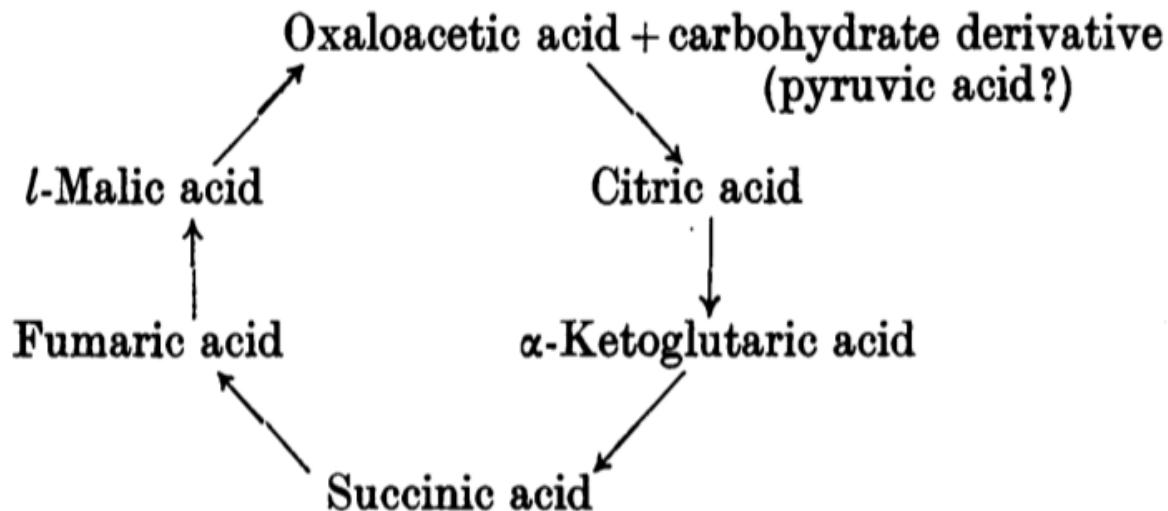
<https://github.com/shuzhao-li/espol-workshop>

updated continuously during this course.

Example of early metabolic studies



Krebs Cycle published 1938



Krebs et al. 1938. Biochem Journal. 32:113

Beadle 1941

Genetic Control of Biochemical Reactions in *Neurospora*

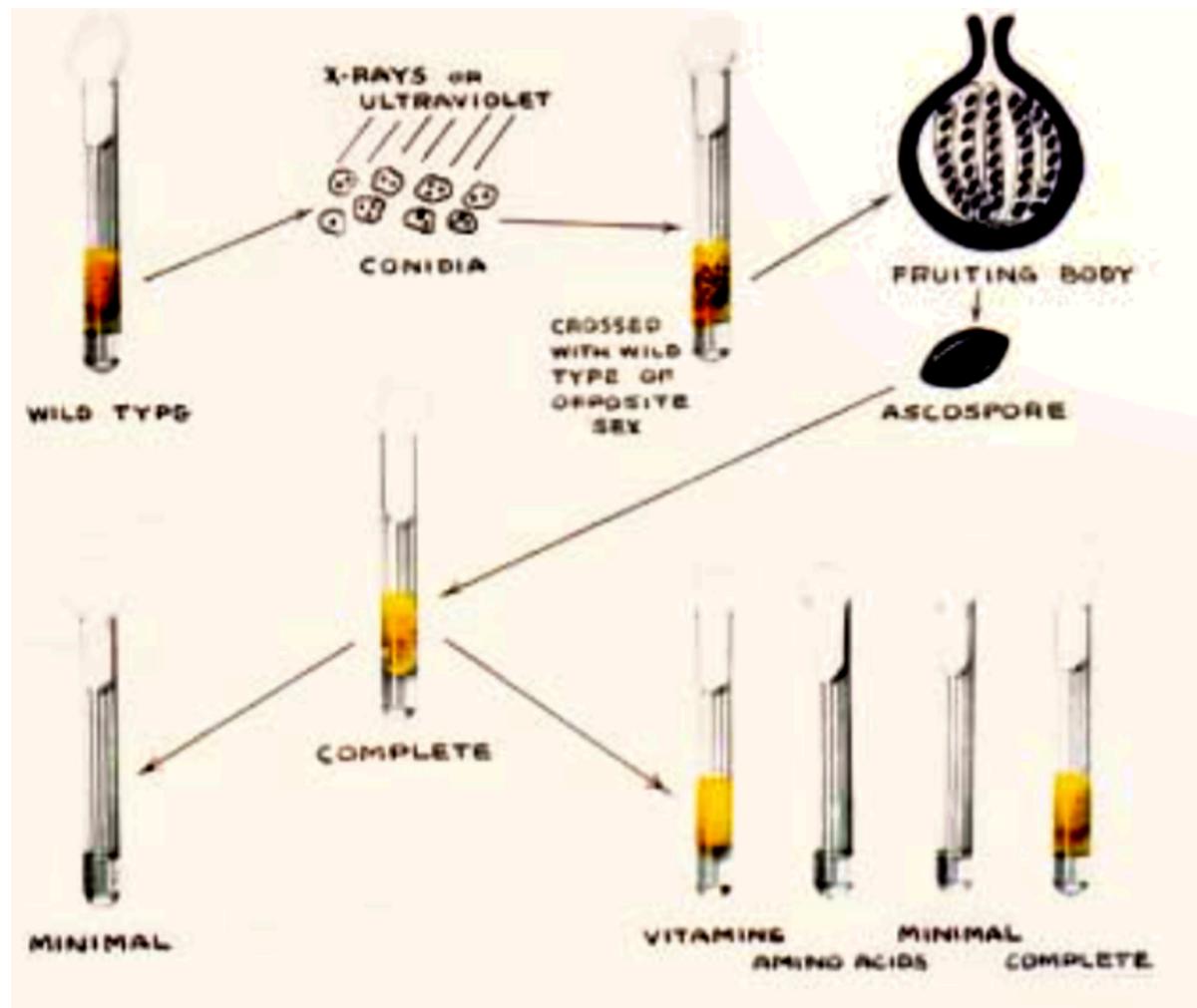


Figure from *Nature Reviews Genetics* 5, 949-954, 2004

Related Nobel Prizes

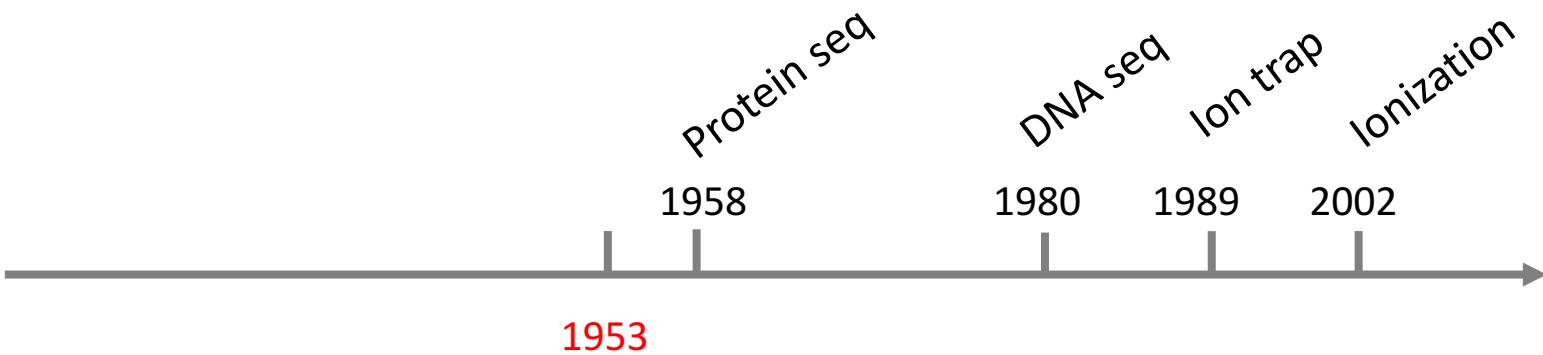
Genetic Control of Biochemical Reactions

1958

1953

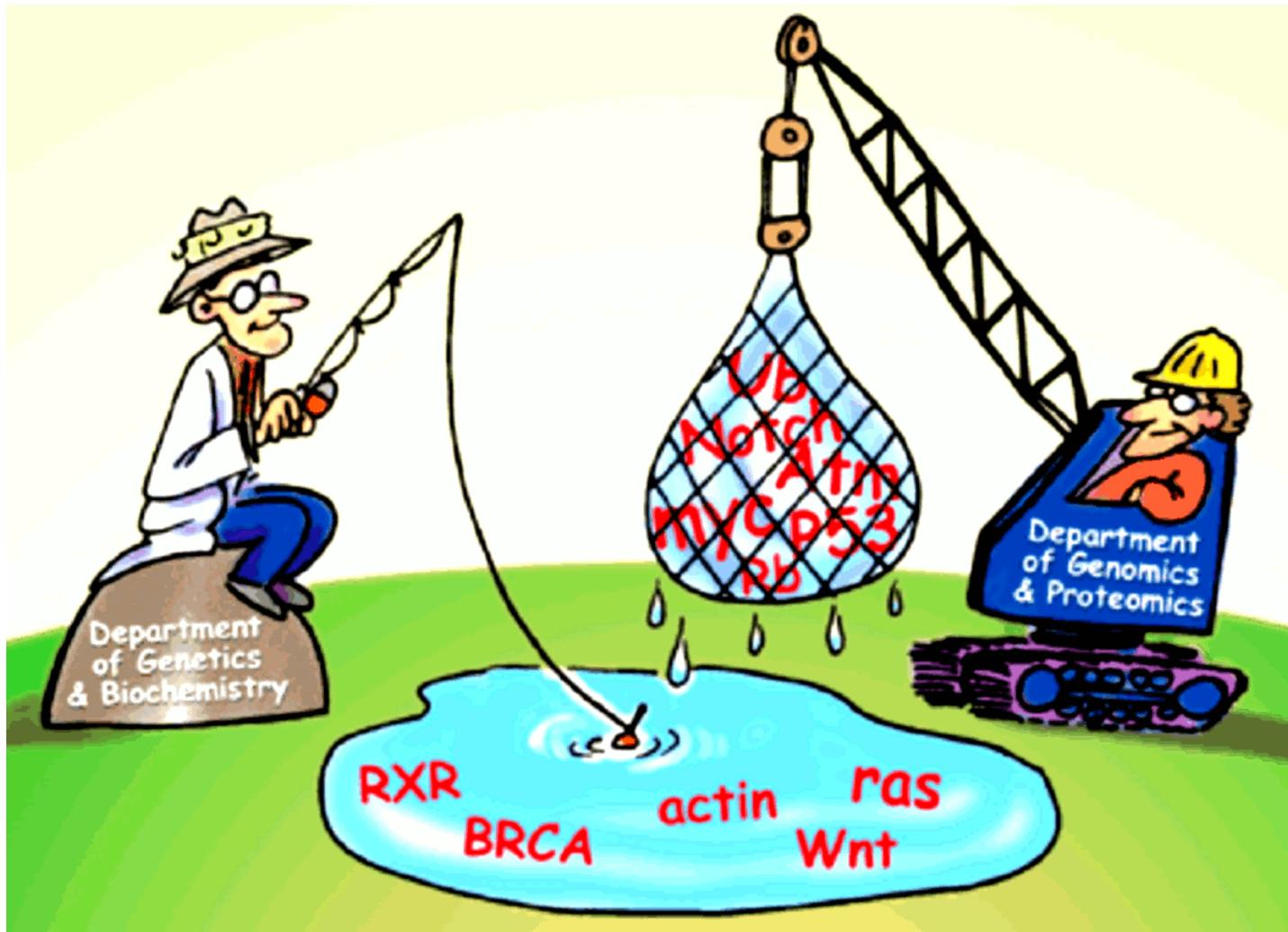
Krebs Cycle

Related Nobel Prizes



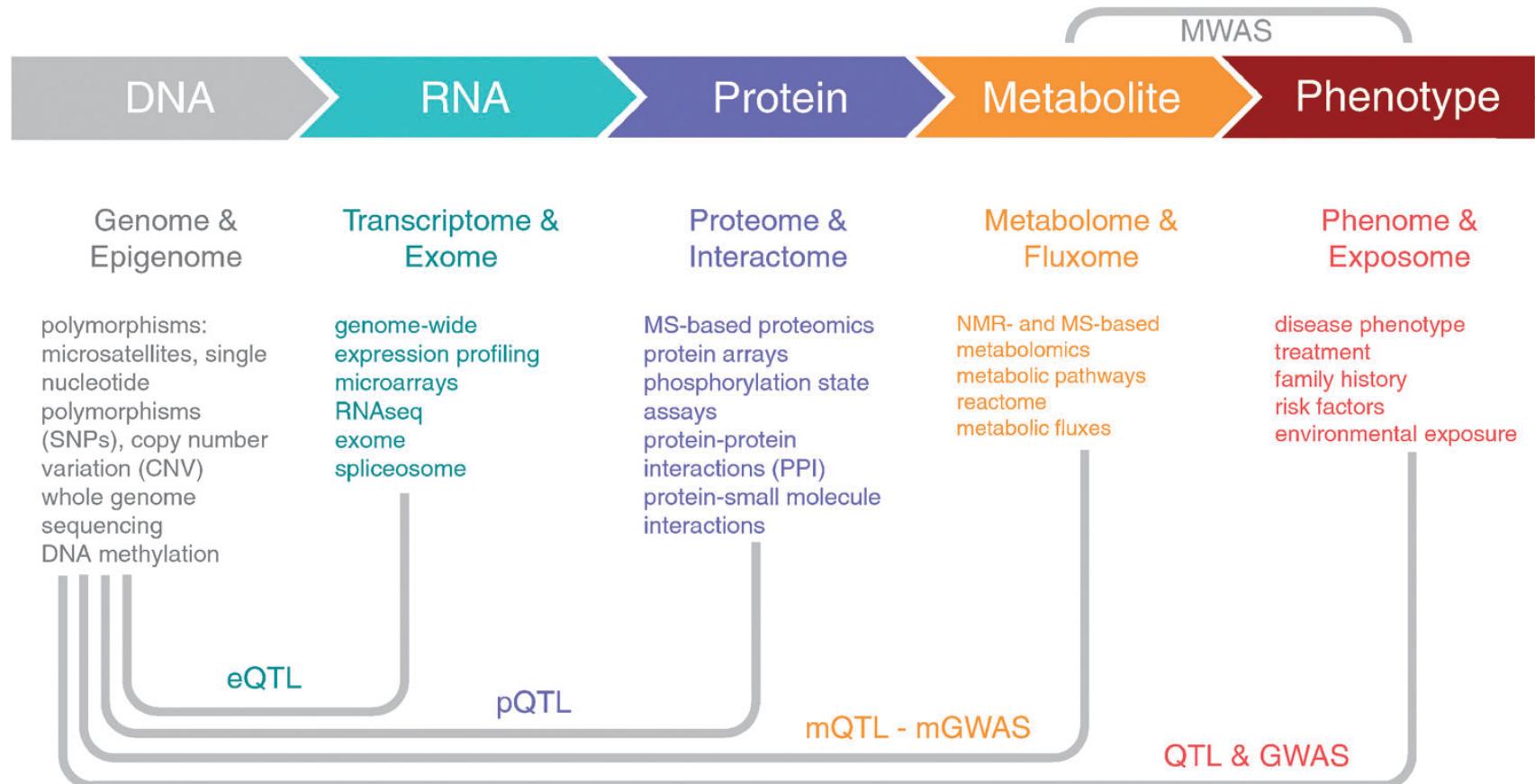
1989 Hans Dehmelt and Wolfgang Paul (ion trap)
2002 John Bennett Fenn (ESI) and Koichi Tanaka (SLD)

When science goes -omics

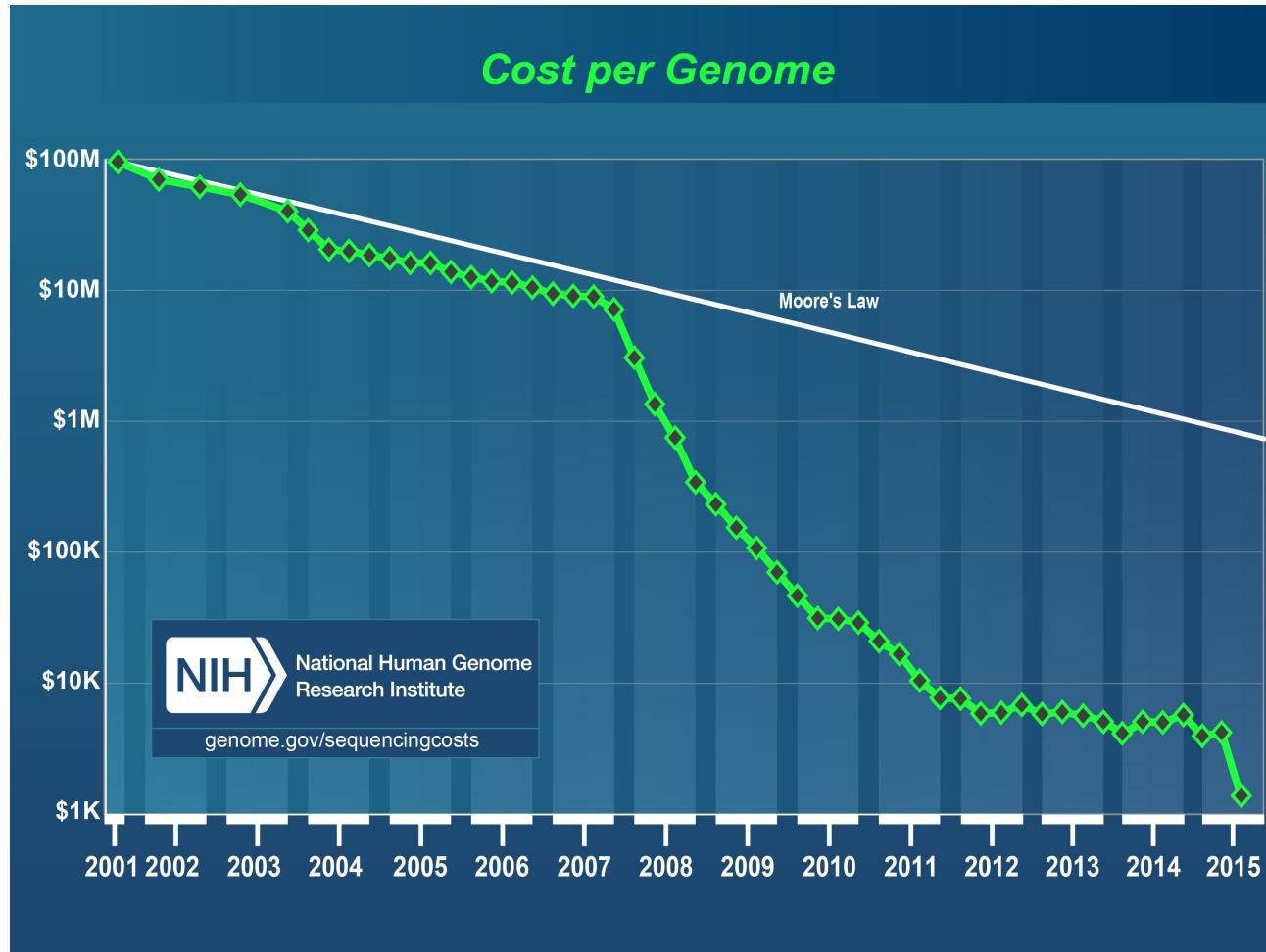


Science (2001) 291:1221

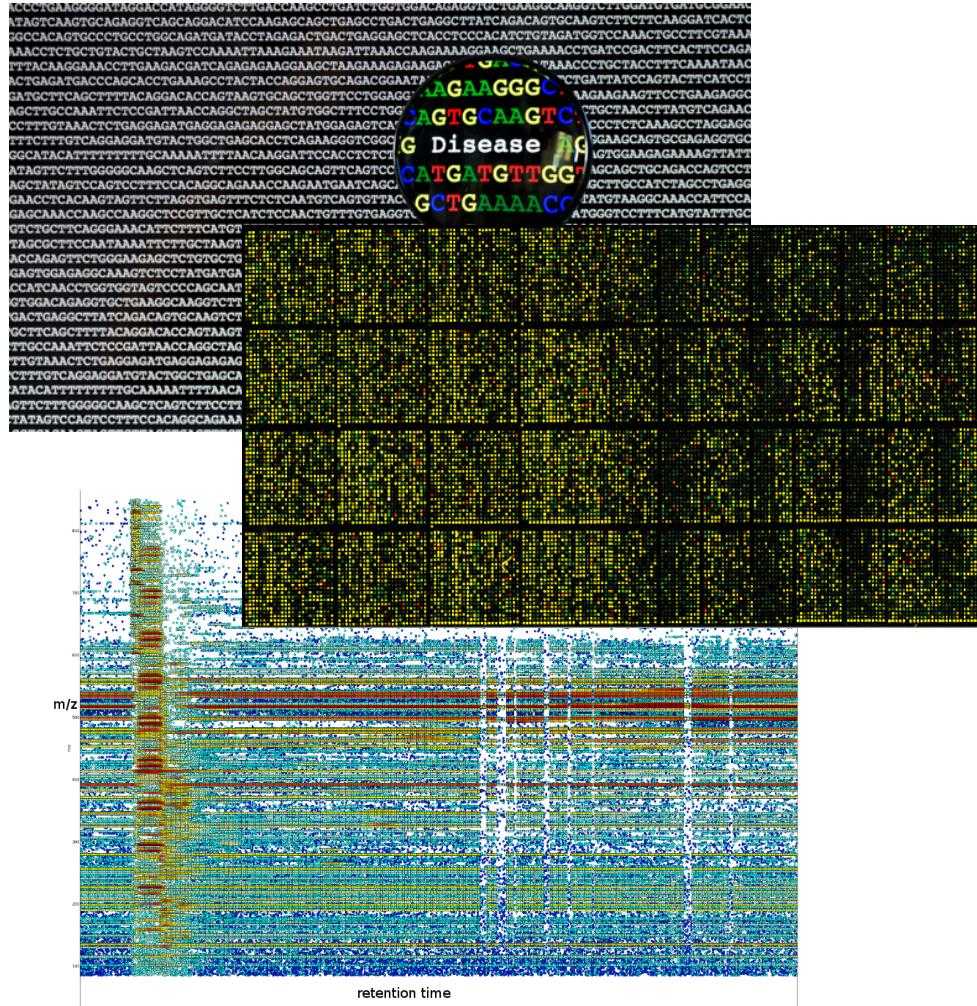
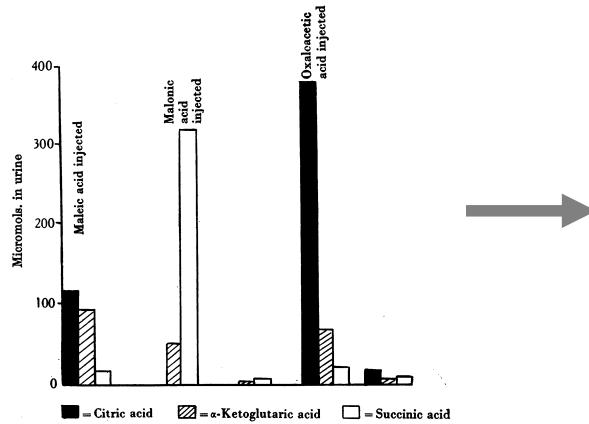
Biomedical -omics data



Economy of data generation (genomics)



The data deluge

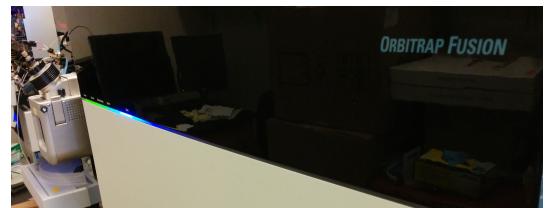
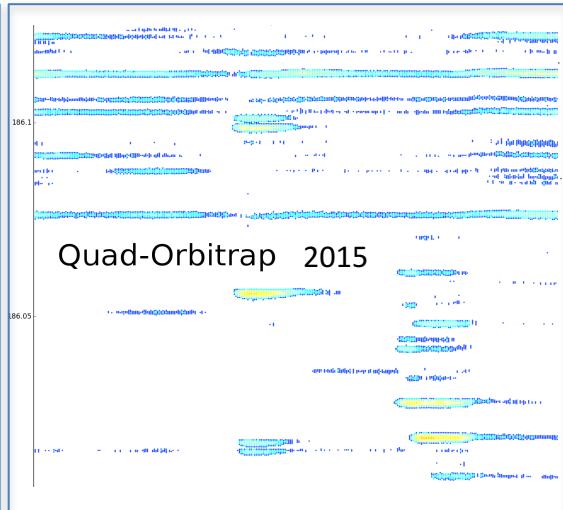
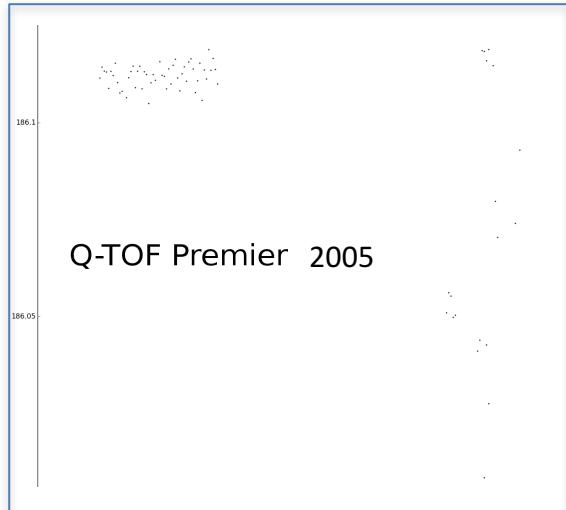
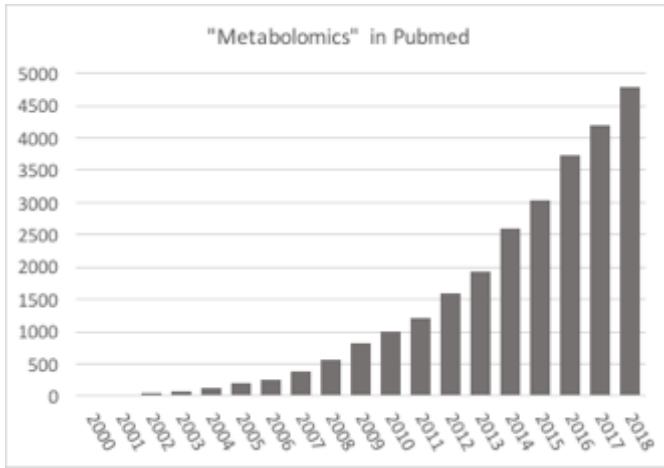


Working with big data

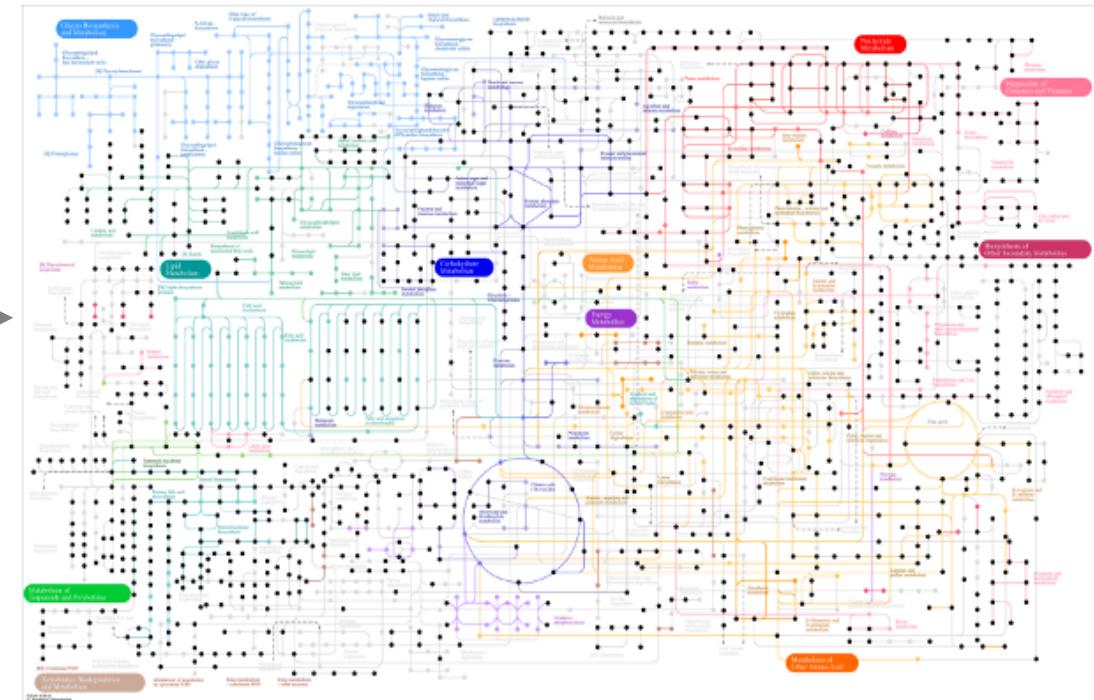
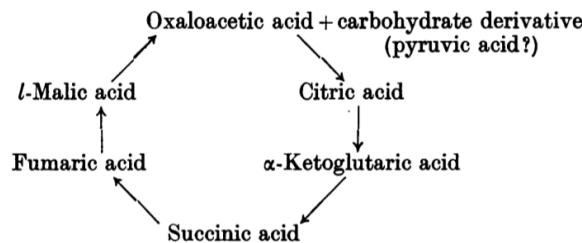
- Global profiling may support many competing hypotheses
- False discovery rate
- Latent structures in data
- Statistical analysis and inference, machine learning
- Combining prior knowledge and domain knowledge
- Integration of different types of data and evidence
- Simulation

Advance of Metabolomics

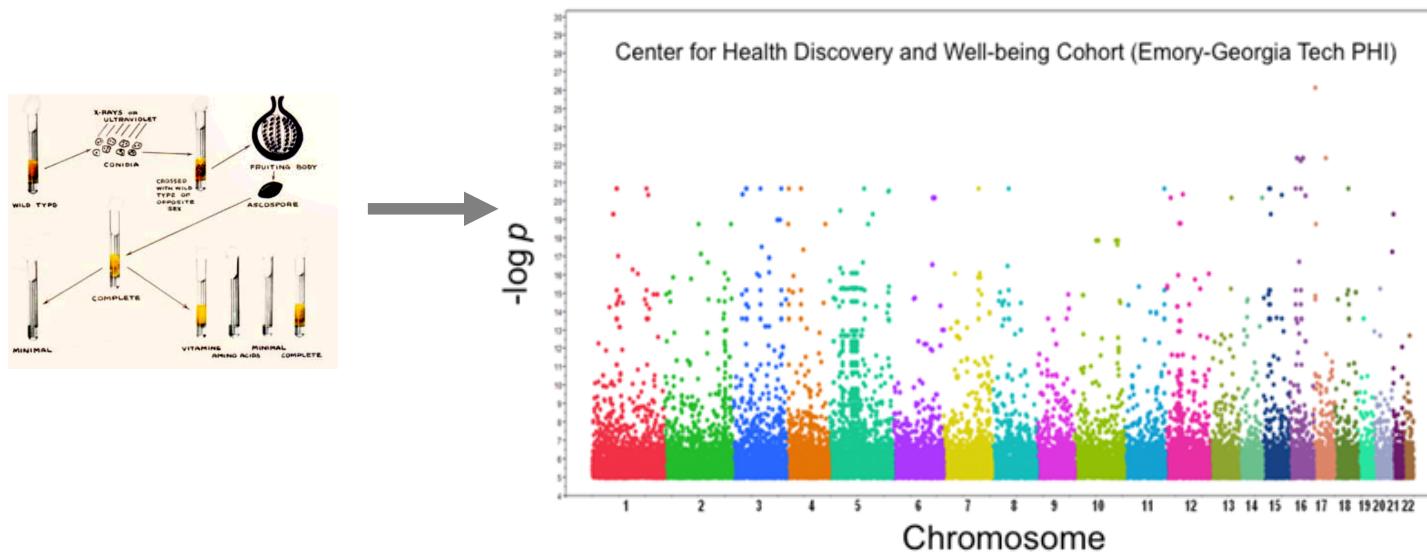
Number of papers



Newer metabolomics cover majority of metabolic pathways

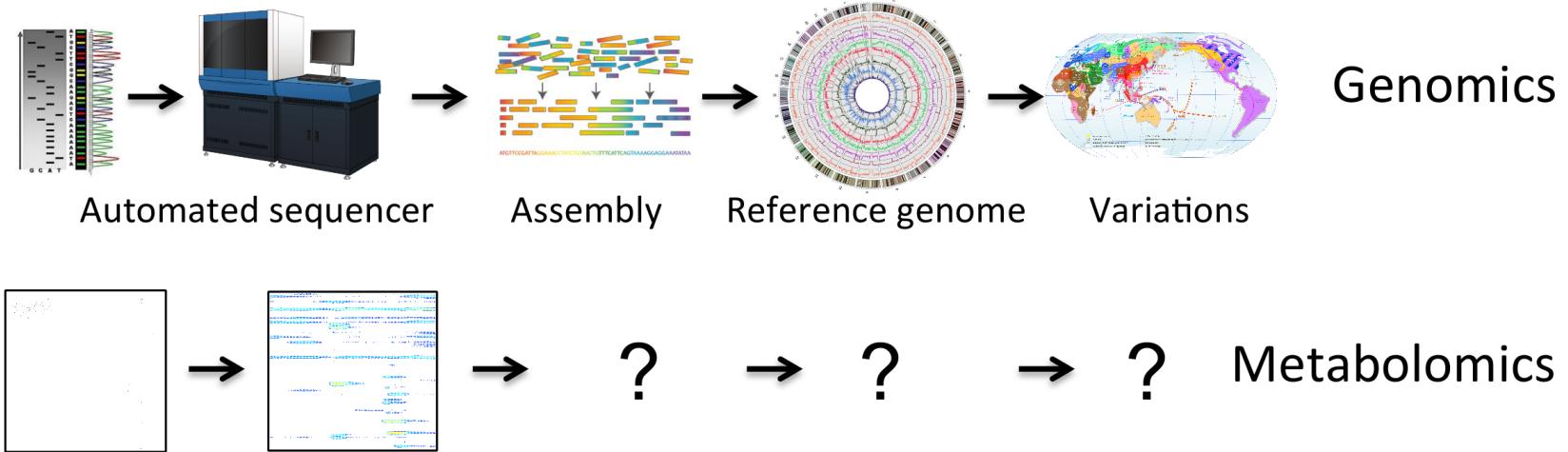


$G \times M$, Metabolome wide association of genetic variations

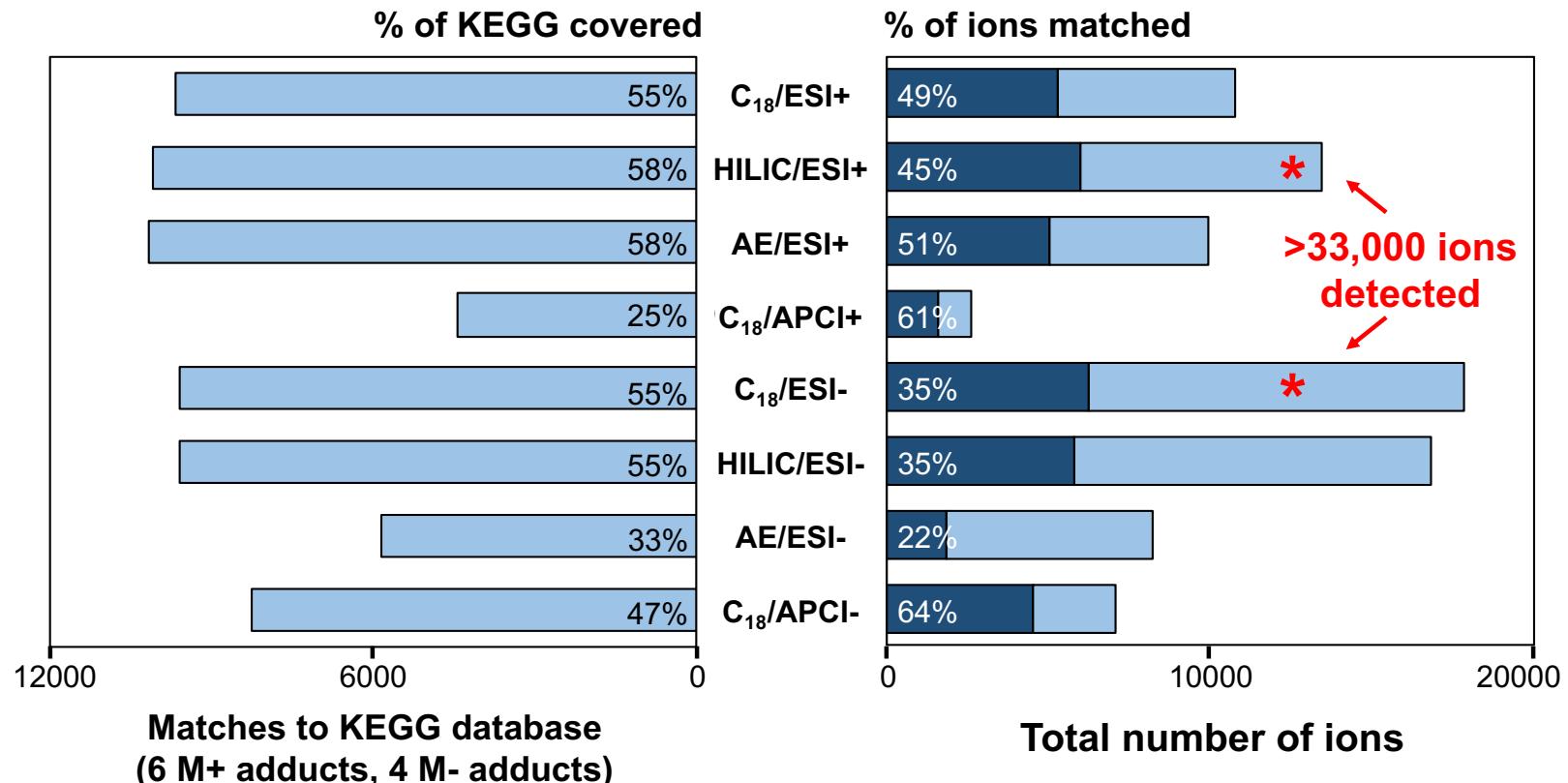


Courtesy: Dean P. Jones and Greg Gibson

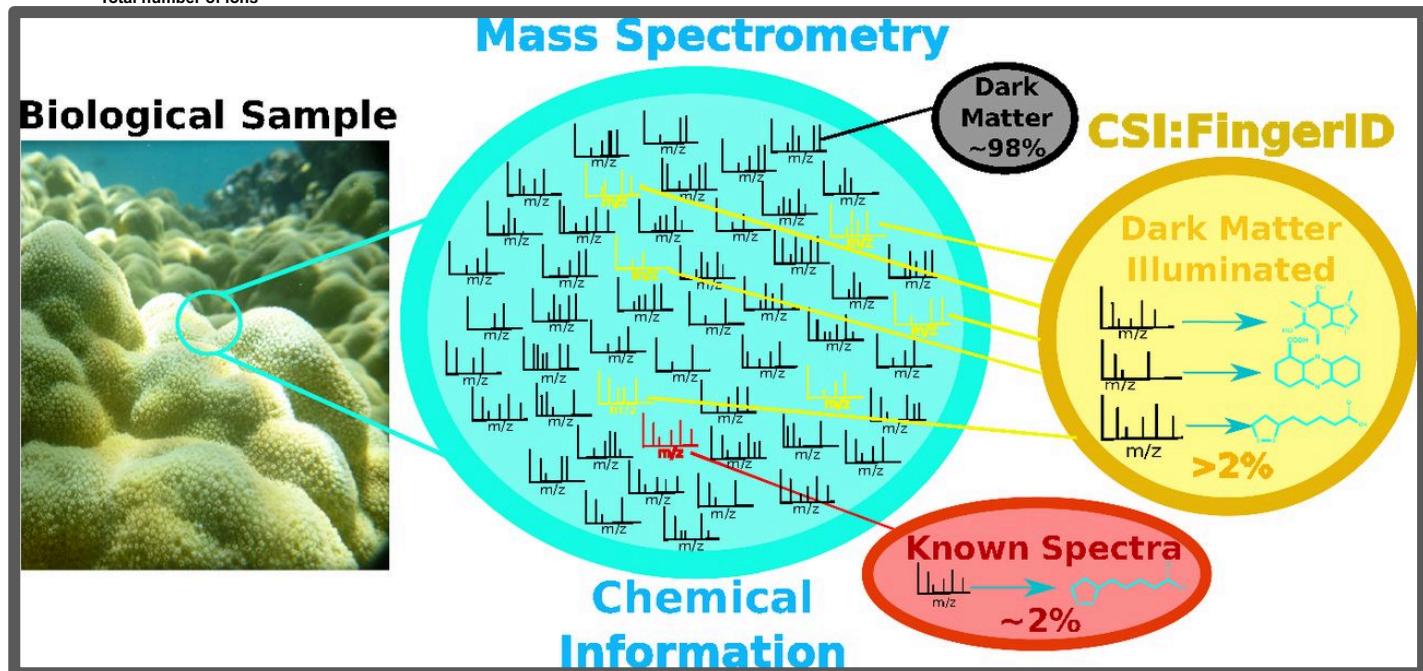
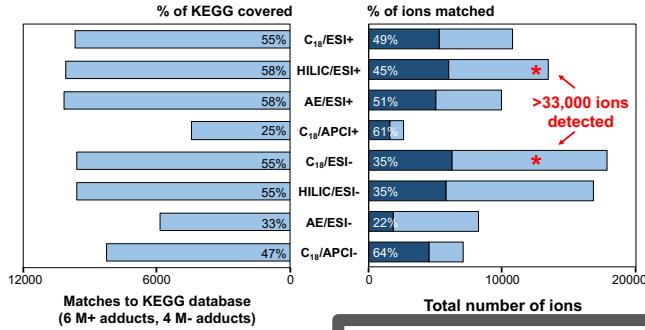
Challenges and Opportunities of Metabolomics



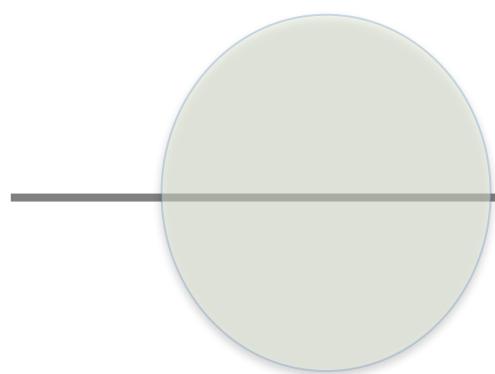
Most metabolomic peaks are unidentified



Dark matters of Metabolomics



Biochemistry 1.0



1953

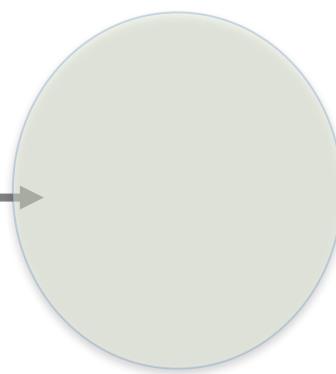
1958

1980

1989

2002

Biochemistry 2.0



Metabolomics is Biochemistry 2.0

- Catalog unidentified chemicals in humans
- Update of metabolic models, reactions and pathways (new maps)
- Signaling and regulatory roles of metabolites and chemicals
- Linking genomes and environmental exposures (exposome), including microbial products, diet and drugs
- Enabling precision medicine with digital health
- Data and computation intensive

NIH Metabolomics Consortium Phase II

<https://metabolomics.info/>

- Compound Identifications
- Informatics tools

Data repositories

Metabolomics workbench

(<https://www.metabolomicsworkbench.org/>)

MetaboLights (EBI

<https://www.ebi.ac.uk/metabolights/>

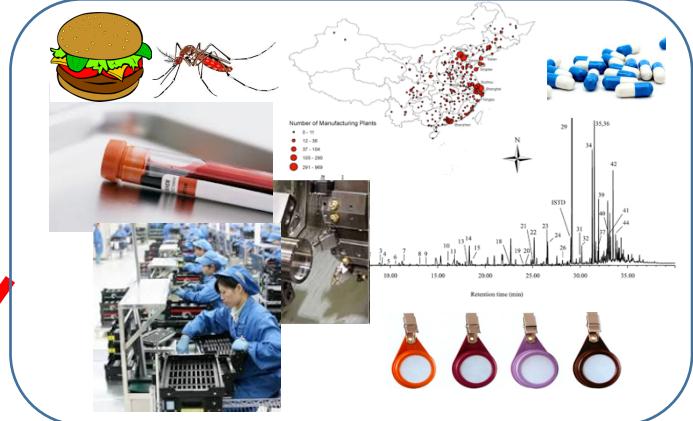
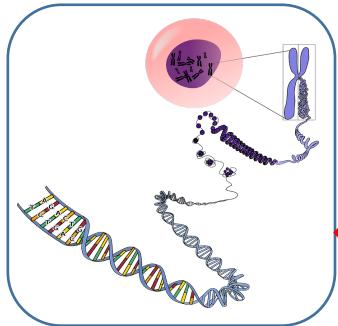
massIVE (overlap proteomics)

<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>

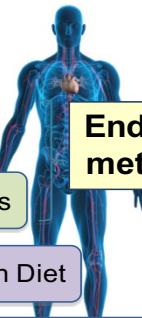
G × M × E

Environment

Genome



Gene function



Core Biological Metabolome

Microbiome-related Chemicals

Non-nutritive Chemicals in Diet

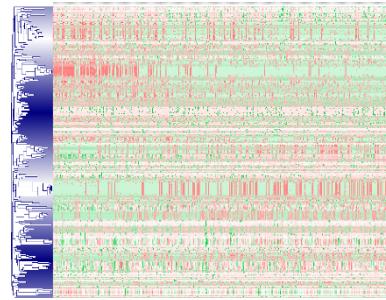
Supplements and Pharmaceuticals

Environmental metabolome

Commercial Products

Environmental Chemicals

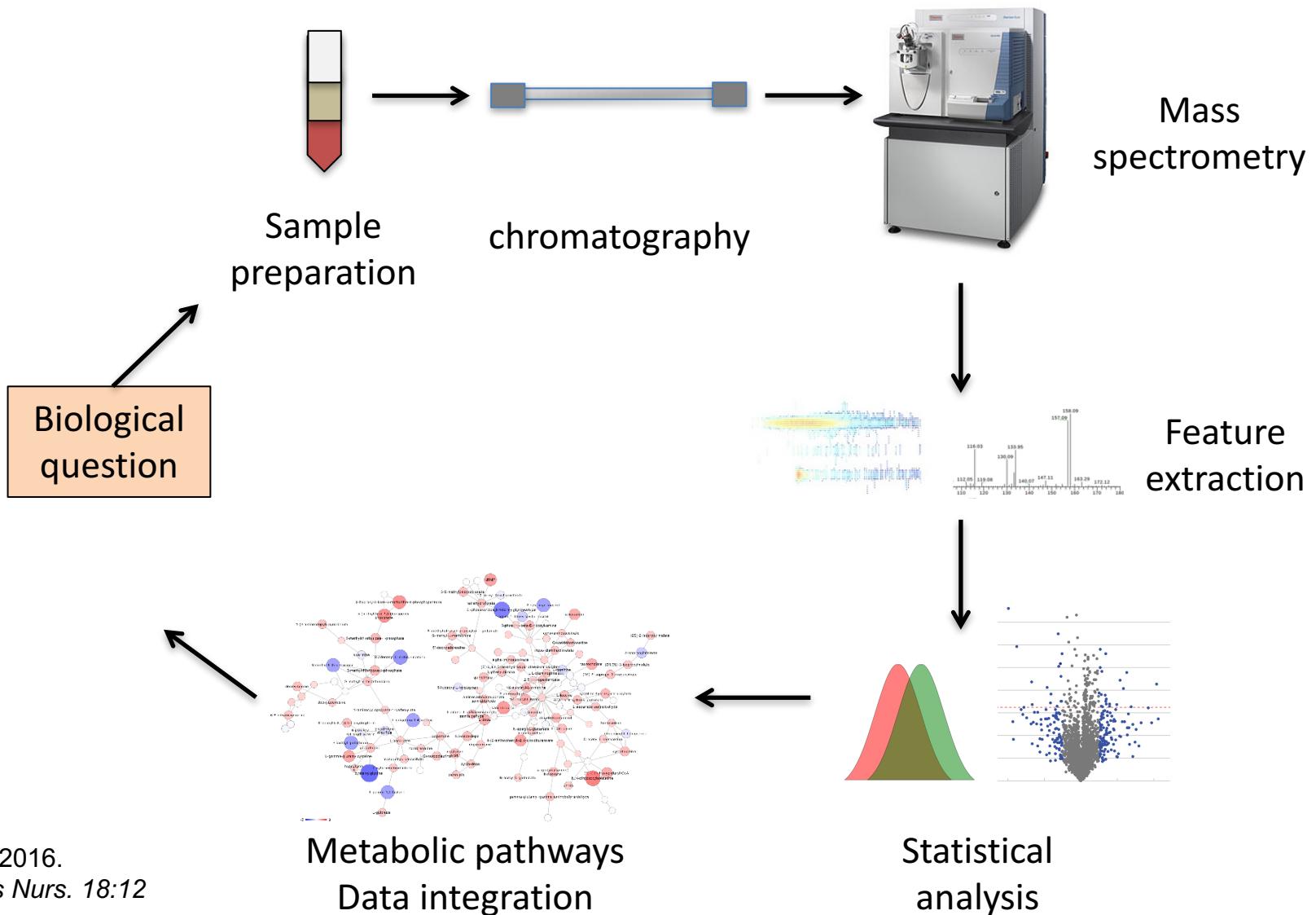
Molecular response



Metabolome

Body burden

Metabolomics workflow (simplified)



Metabolomics technologies / platforms

- NMR vs Mass spectrometry
- Targeted vs untargeted
- LC vs GC
- MS vs MSⁿ
- IM, EC, ...
- Isotope tracing

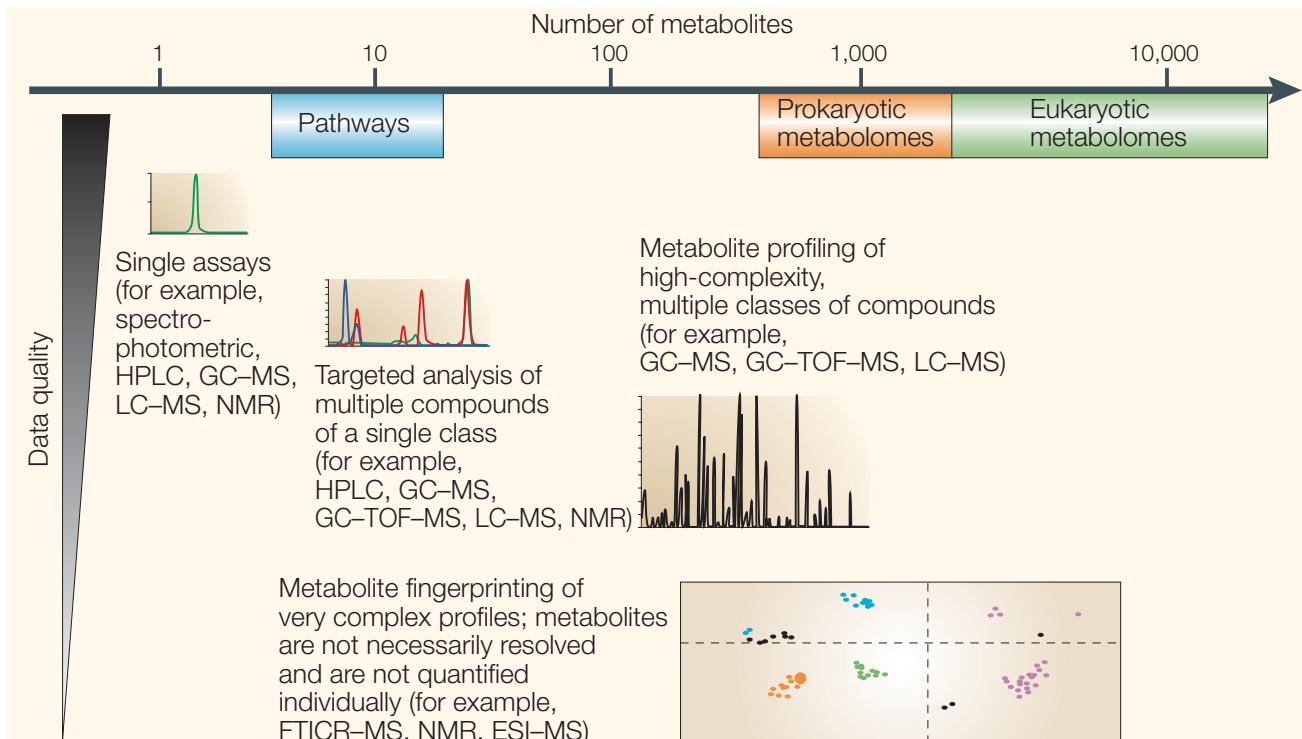


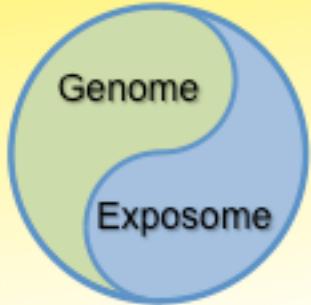
Figure 1 | The trade-off between metabolic coverage and the quality of metabolic analysis.

Fernie, A.R., Trethewey, R.N., Krotzky, A.J. and Willmitzer, L., 2004. Metabolite profiling: from diagnostics to systems biology. *Nature reviews molecular cell biology*, 5(9), p.763.

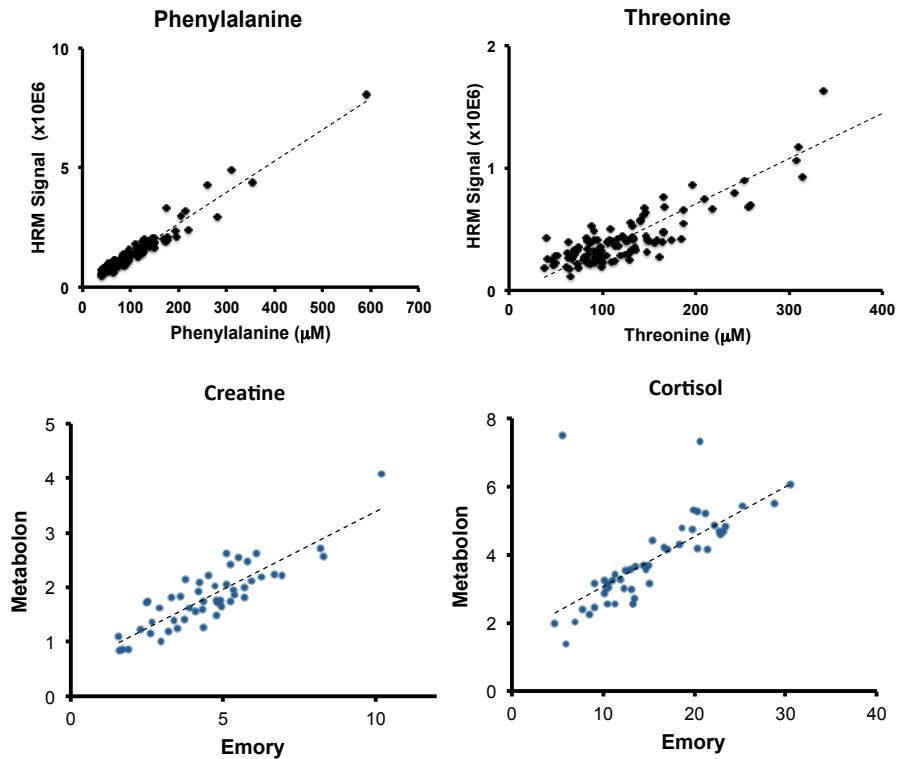
Table 1 | A comparison of different metabolomics technologies

Technology	Advantages	Disadvantages	Refs
NMR spectroscopy	<ul style="list-style-type: none"> Quantitative Non-destructive Fast (2–3 min per sample) Requires no derivatization Requires no separation Detects most organic classes Allows identification of novel chemicals Most spectral features are identifiable Robust, mature technology Can be used for metabolite imaging (fMRI or MRS) Can be fully automated Compatible with liquids and solids Long instrument lifetime (over 20 years) 	<ul style="list-style-type: none"> Not sensitive ($LOD = 5 \mu\text{M}$) High start-up cost (>US\$1 million) Large instrument footprint Cannot detect or identify salts and inorganic ions Cannot detect non-protonated compounds Requires larger sample volumes (0.1–0.5 mL) 	17,18,35
GC-MS	<ul style="list-style-type: none"> Robust, mature technology Modest start-up cost (~\$150,000) Quantitative (with calibration) Modest sample volume (0.1–0.2 mL) Good sensitivity ($LOD = 0.5 \mu\text{M}$) Large body of software and databases for metabolite identification Detects most organic and some inorganic molecules Excellent separation reproducibility Many spectral features are identifiable Can be mostly automated Compatible with gases and liquids 	<ul style="list-style-type: none"> Destructive (sample not recoverable) Requires sample derivatization Requires separation Slow (20–40 min per sample) Cannot be used in imaging Not compatible with solids Novel compound identification is difficult 	18–20
LC-MS	<ul style="list-style-type: none"> Superb sensitivity ($LOD = 0.5 \text{ nM}$) Very flexible technology Detects most organic and some inorganic molecules Small sample volumes (10–100 μL) Can be used in metabolite imaging (MALDI or DESI) Can be done without separation (direct injection) Has the potential to detect the largest portion of metabolome Can be mostly automated Compatible with solids and liquids 	<ul style="list-style-type: none"> Destructive (sample not recoverable) Not very quantitative Higher start-up cost (>\$300,000) Slow (15–40 min per sample) Usually requires separation Poor separation resolution and lower reproducibility versus GC-MS Less-robust instrumentation than NMR or GC-MS Not compatible with gases Most spectral features are not yet identifiable Novel compound identification is difficult Short instrument lifetime (<9 years) 	19,20,33,38

DESI, desorption electrospray ionization; fMRI, functional MRI; GC-MS, gas chromatography mass spectrometry; LOD, limit of detection; LC-MS, liquid chromatography mass spectrometry; MRS, magnetic resonance spectroscopy; MALDI, matrix-assisted laser desorption/ionization; NMR, nuclear magnetic resonance.



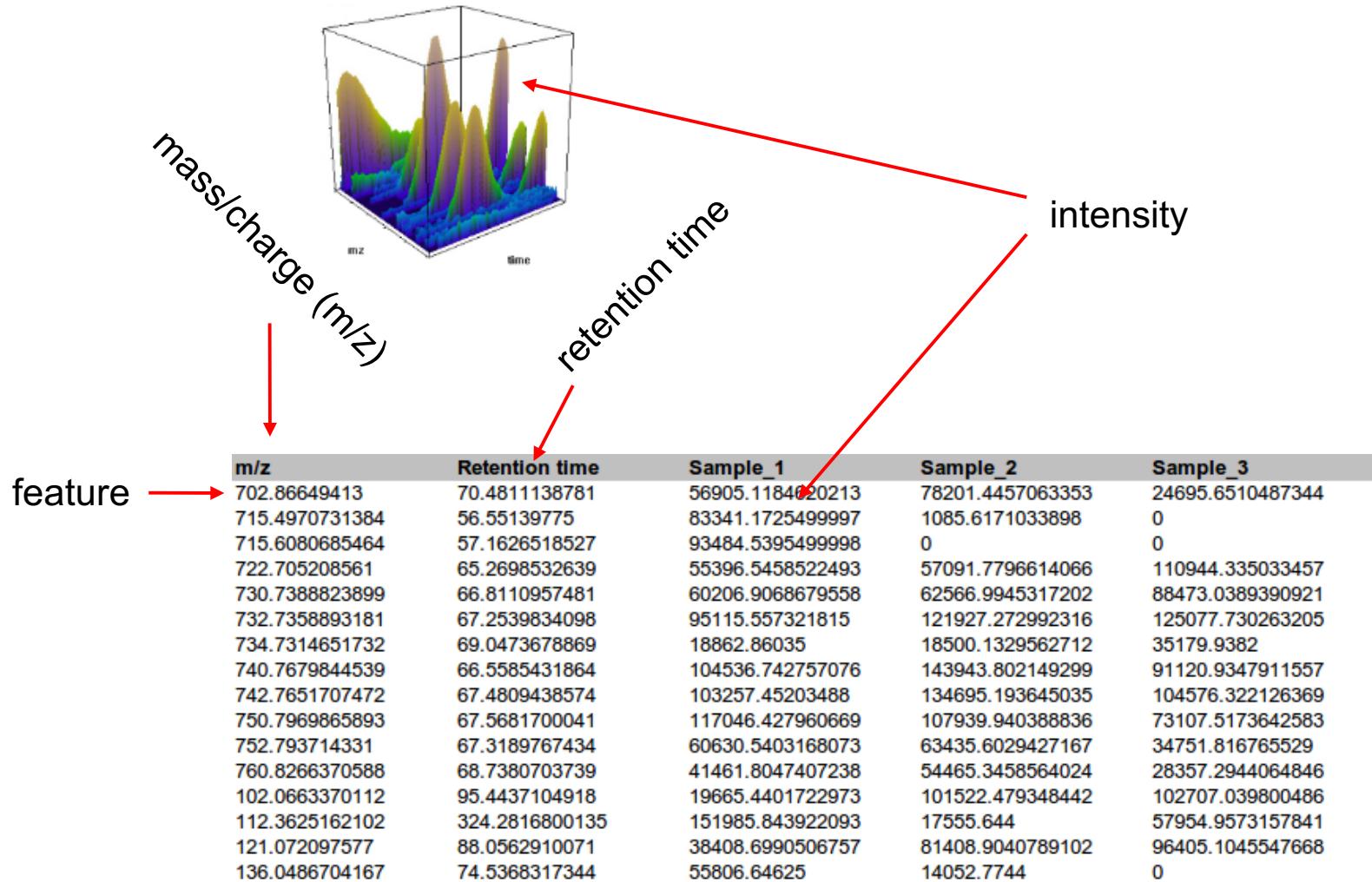
Accurate quantification in high-throughput mode shown by cross-laboratory comparisons



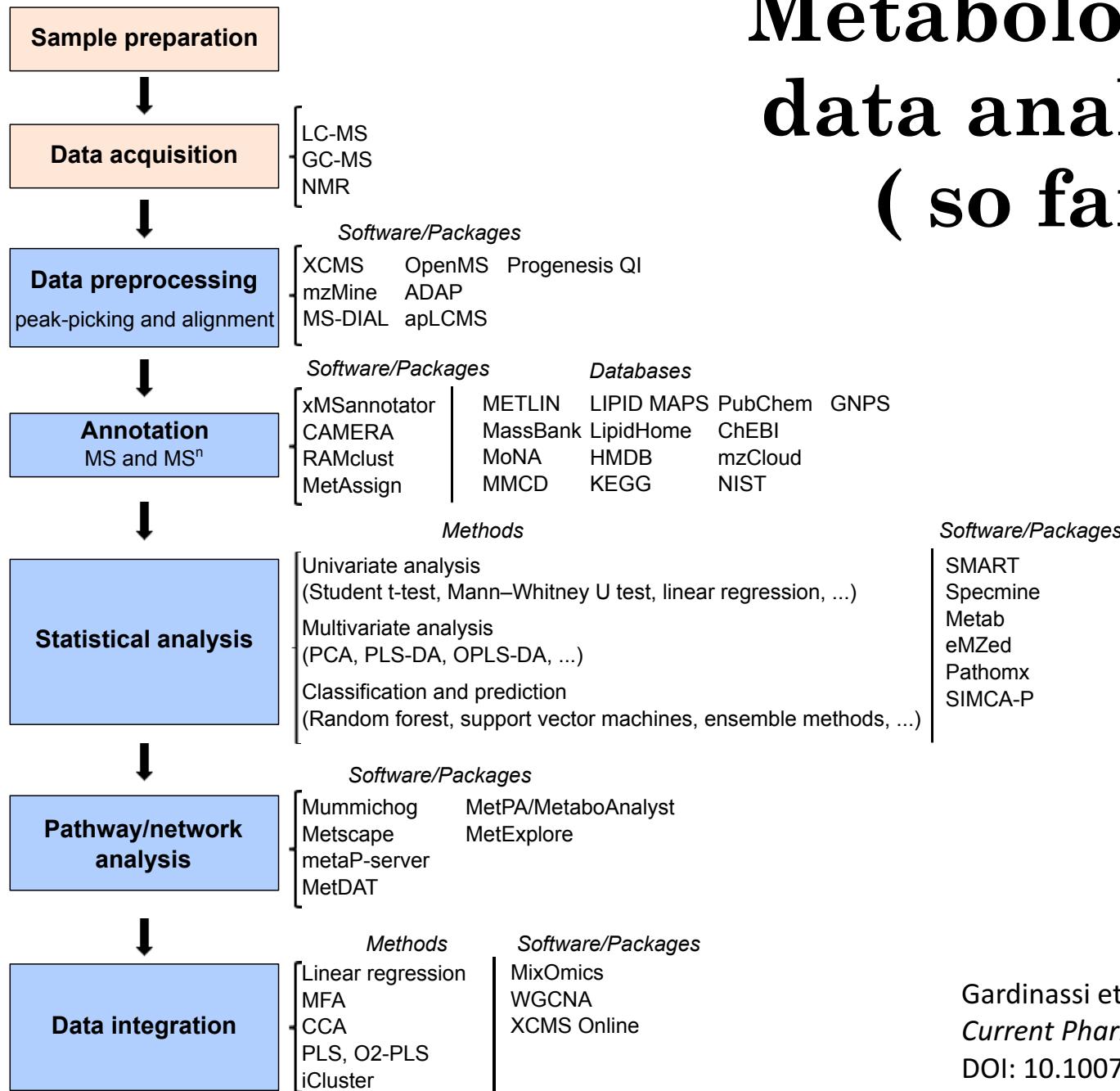
HRM metabolite quantification in 30 orphan samples

Identity	Mean \pm SD (μM)	HMDB (μM)
Arginine	148 ± 39	60 to 140
Glycine	280 ± 62	212-329
Histidine	100 ± 12	75 to 143
Ornithine	83 ± 28	54 to 94
Phenylalanine	131 ± 18	48 to 88
Threonine	136 ± 22	102 to 260
Tryptophan	56 ± 7	44 to 78
Tyrosine	84 ± 23	54 to 143
Glucose	4310 ± 1153	3900 to 6100
Kynurenine	2.0 ± 0.4	1.4 to 2.4
Carnitine	52 ± 9	30 to 57
Creatinine	93 ± 13	59 to 109
Creatine	16 ± 8	8.4 to 65

Metabolomics data (LC-MS)



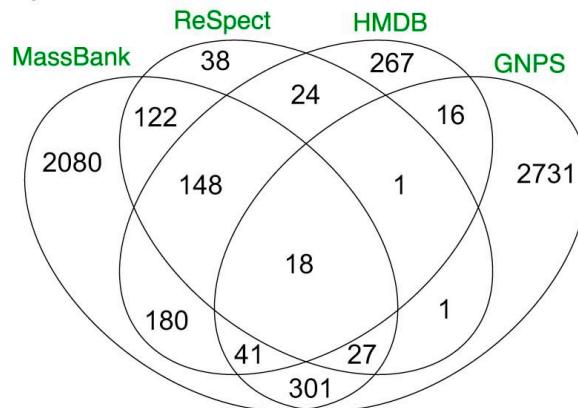
Metabolomics data analysis (so far)



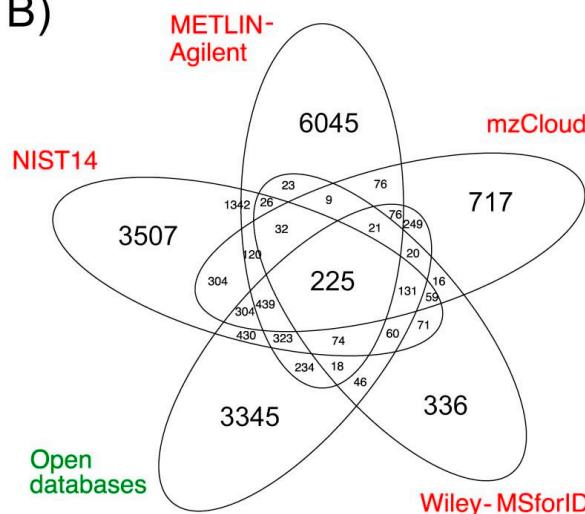
Gardinassi et al (2017)
Current Pharmacology Reports,
DOI: 10.1007/s40495-017-0107-0.

Mass Databases (2016)

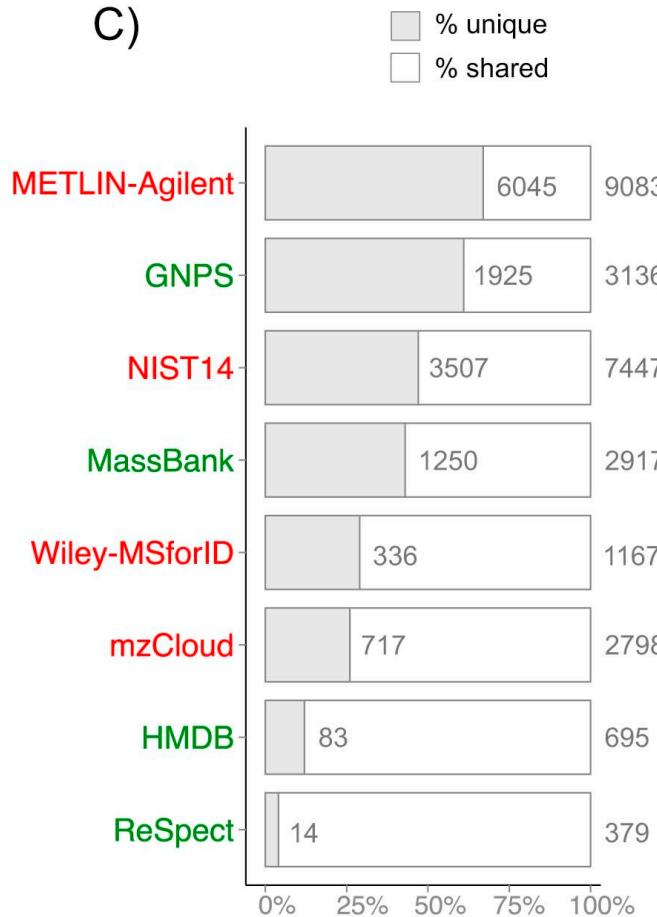
A)



B)



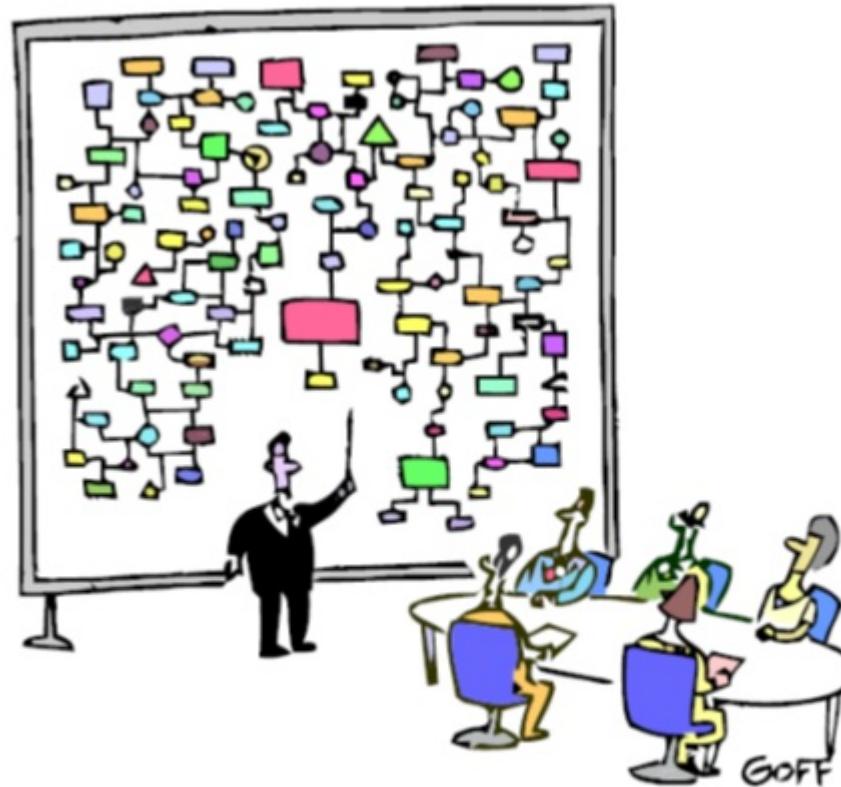
C)



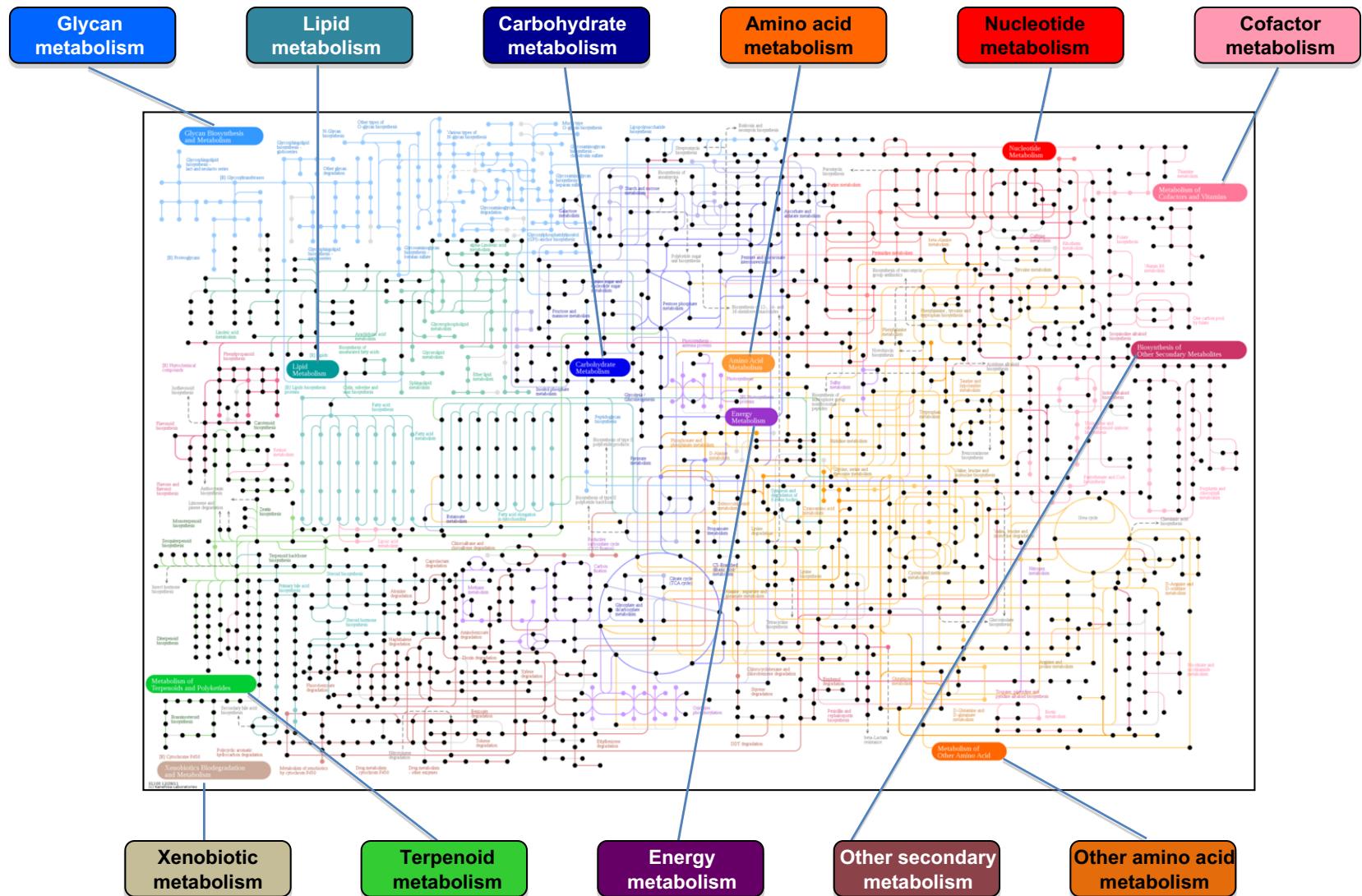
(A) Venn diagram showing the overlap between open mass spectral databases (HMDB, MassBank, GNPS, and ReSpect). (B) Venn diagram showing the overlap between five commercial databases (Agilent METLIN PMD, mzCloud, NIST 14, and Wiley MS) and open databases described in A. (C) Number and percentage of unique and shared compounds (i.e., InChIKey) with MS^n ($n \geq 2$) data in each database in relation to all eight resources.

Vinaixa, M., Schymanski, E.L., Neumann, S., Navarro, M., Salek, R.M. and Yanes, O., 2016. Mass spectral databases for LC/MS-and GC/MS-based metabolomics: state of the field and future prospects. *TrAC Trends in Analytical Chemistry*, 78, pp.23-35.

Lot of pathways are known



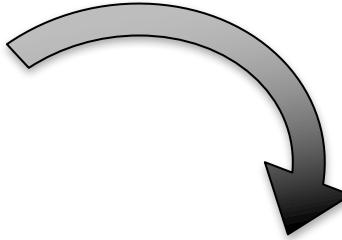
"And that's why we need a computer."



Towards new metabolic maps



1482

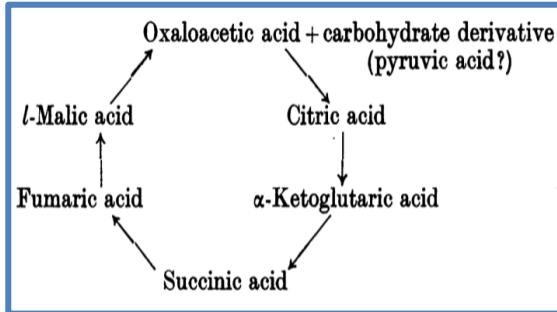
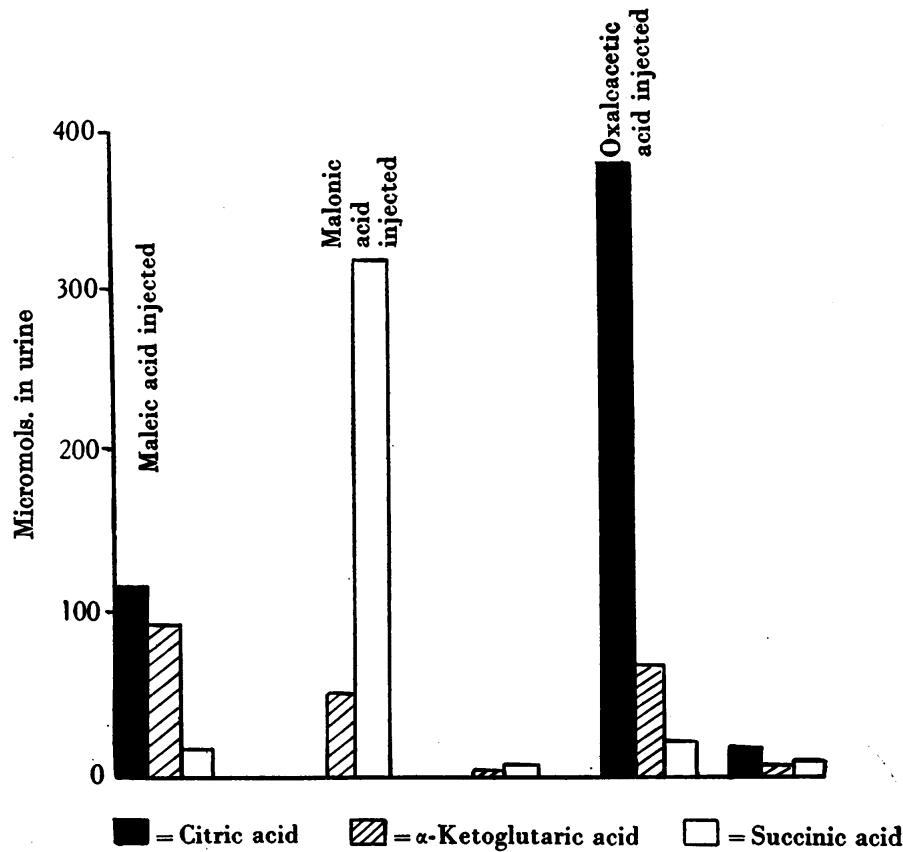
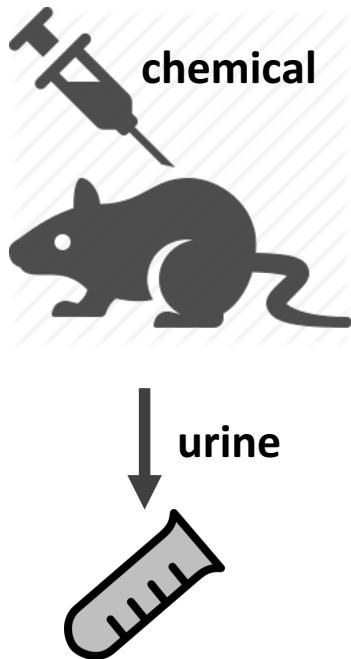


Political Map of the World, January 2015



2015

Charting metabolic pathways

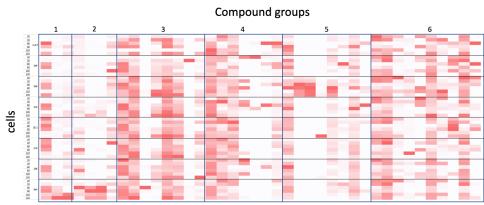


Krebs et al. 1938.
Biochem Journal. 32:113

Charting metabolic pathways (2019)

Unpublished data removed

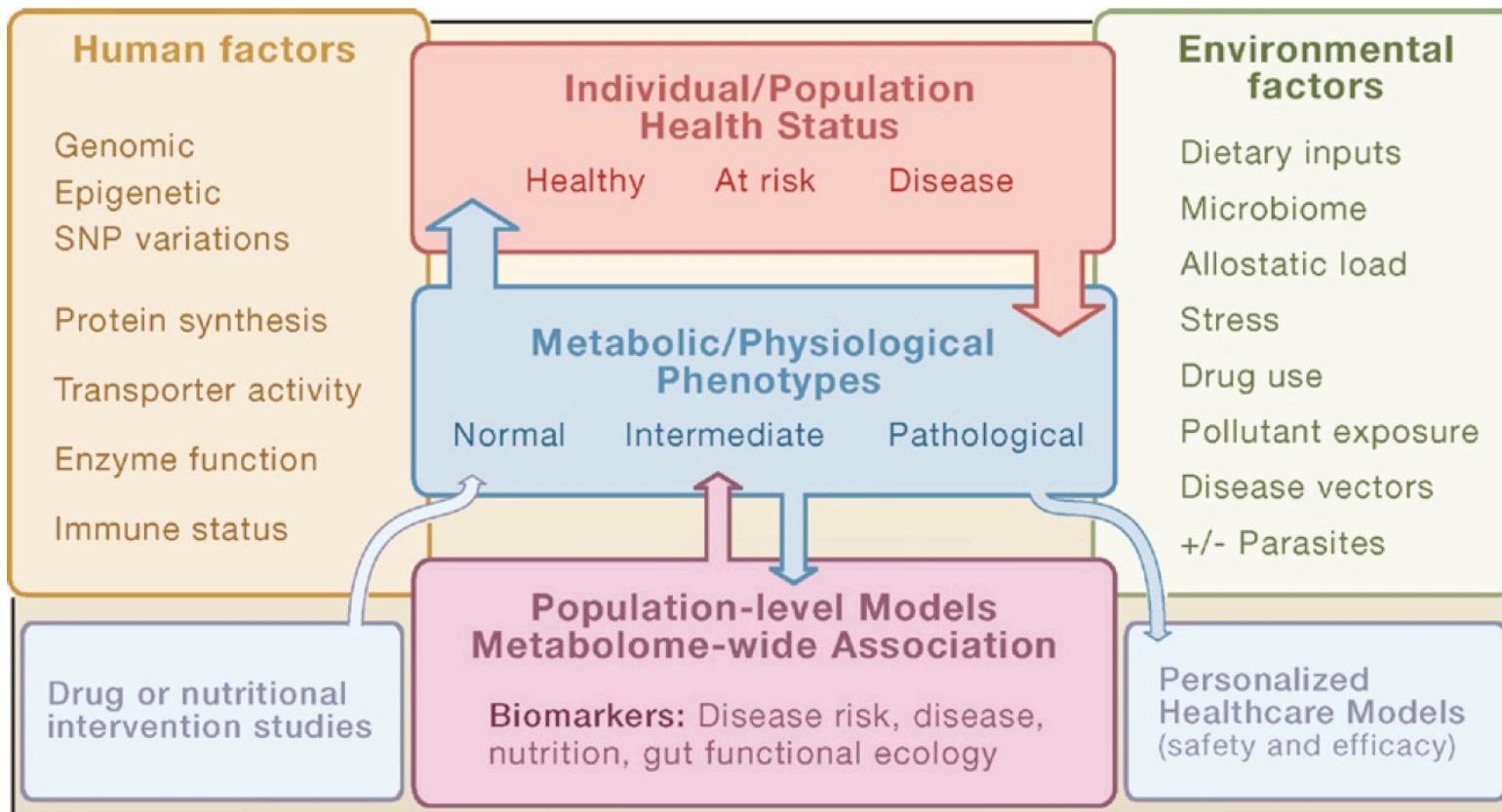
Charting metabolic pathways



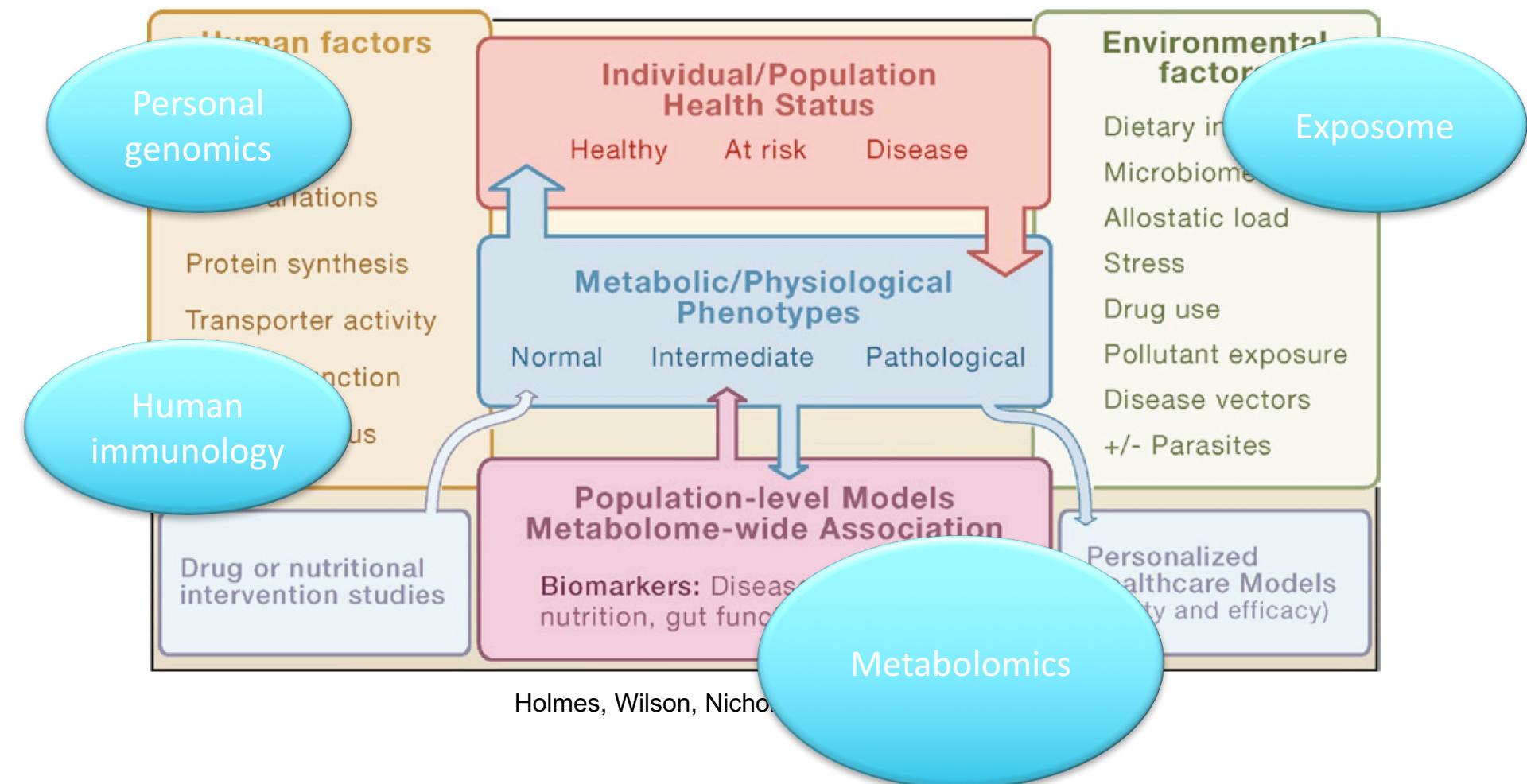
- Capturing unknown; tandem mass spec; isotope tracing in cells and whole organisms
- Combination with genetic tools
- Domain applications, tumors, microbiome, immune cells
- Azimuth database; Interfacing with computational reconstruction of biochemical networks

Modeling systems medicine

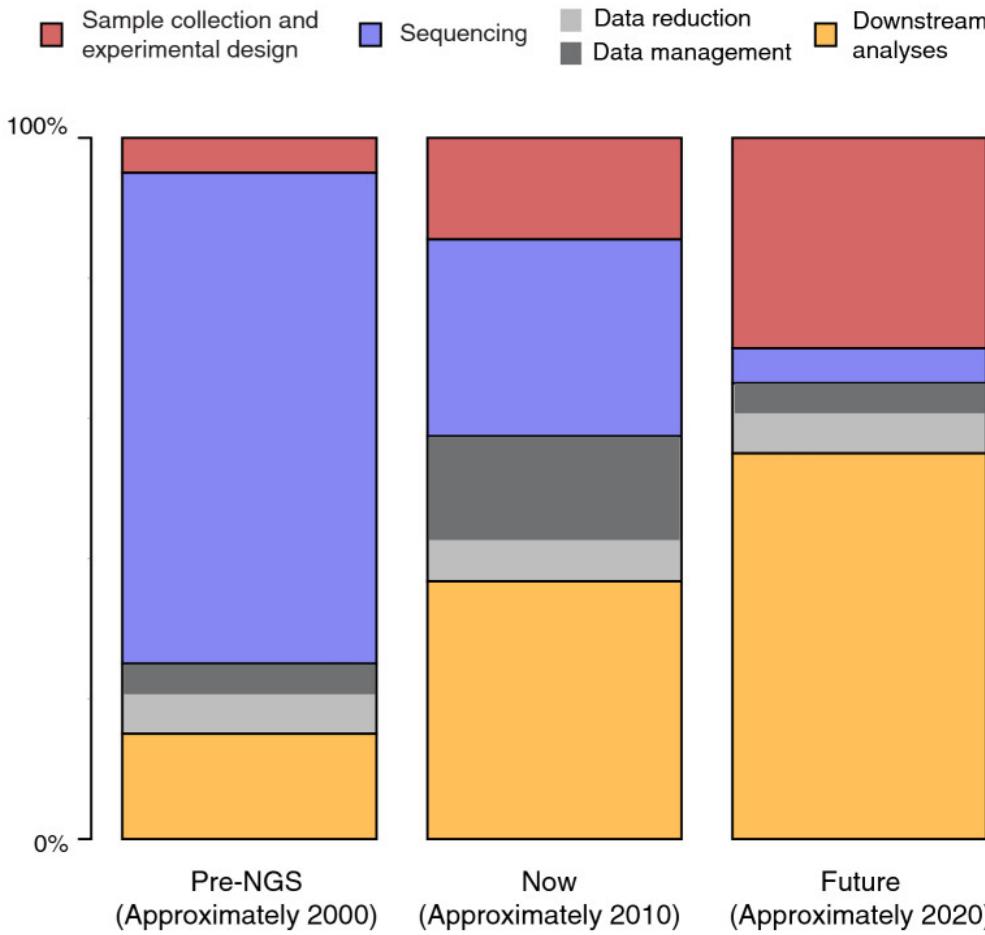
Holmes, Wilson, Nicholson (2008). Cell 134:714



What has changed (2019)



The real cost of omics



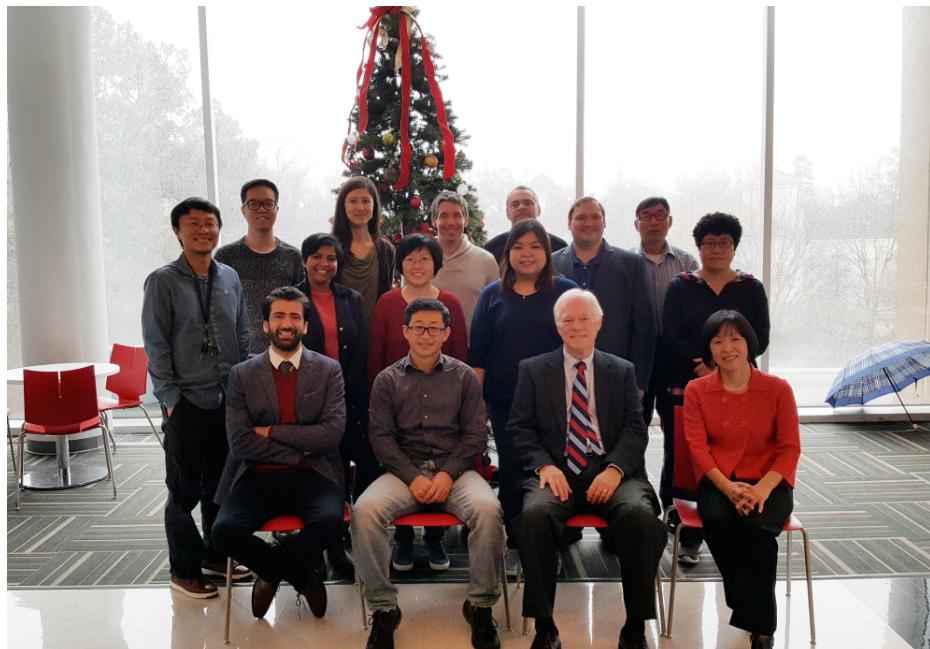
Objectives

- Understanding of experimental design and statistics
- Using Jupyter notebooks for data analysis
- Web tools such as XCMS Online, MetaboAnalyst and Mummichog server
- Metabolite reporting standards
- Discussion of research projects

Acknowledgement

CHDS

Barbara Cohn
Piera Cirrilo
Nickilou Krigbaum
Stanford University
Bali Pulendran
University of Colorado
Myron Levin
Adriana Weinberg



Emory University

Rafi Ahmed
Mark Mulligan (NYU)
Nadine Rousphael
Mary Galinski
Judith Fridovich-Keil
Taylor Fischer
Miriam Vos

Dean P. Jones
Young-Mi Go
Douglas Walker (MSSM)
ViLinh Tran
Bill Liang
Luiz Gardinassi
Ken Liu
M. Ryan Smith
Tianwei Yu
Edward Morgan
Choon Myung Lee
Grant Singer

Funding support from US National Institutes of Health U19 AI090023, HHSN272201300018I, UH2 AI132345, U01 CA235493, U2C ES030163, P30ES019776, U2C ES026560, P50 ES026071, R01 DK107900, R01 AI121252, R01 GM124061, and EPA (83615301)