Multi-MS-Omics Data Integration

ASMS Bioinformatics MS Interest Group

Wednesday Evening Workshop

Samuel Payne & Isabell Bludau

Audience survey:
https://www.surveymonkey.com/r/SRWK2B8

Why “Multi-MS-Omics Data Integration”?

- Individual ‘omics’ disciplines can reveal valuable biological insights
- Cellular processes are more completely described by the diversity and **interplay of all different types of molecules**
- As technologies in proteomics, metabolomics and lipidomics improve, it is critical to **remain connected as a community**

➪ **Discuss approaches to integrate and benefit from multi-omics data**

Why – Part II

- Clinical multi-omics investigations
  - UDN
  - MotrPac
  - Cancer – CPTAC/TCGA

- Synthetic Biology
  - Pathway flux optimization

Outline

- Current status of **proteomics**, **metabolomics** and **lipidomics** research and data analysis with their specific benefits and limitations
- Current strategies to perform **multi-omics data integration** to increase biological insights
- Novel analysis workflows to investigate **cross-omics interaction networks**
- **Audience quiz & panel discussion** (prepare yourself for active contribution ☺)

Invited experts:

- Hannes Röst − proteomics & metabolomics
- Jeremy Kolmel − lipidomics
- Ilaria Piazza − protein-metabolite interactions

Fill out our survey

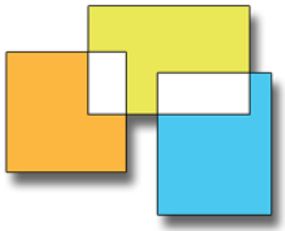
Audience survey:
https://www.surveymonkey.com/r/SRWK2B8

Personalized profiling & Multi Omics integration

Hannes Röst, PhD

University of Toronto

2018-07-06

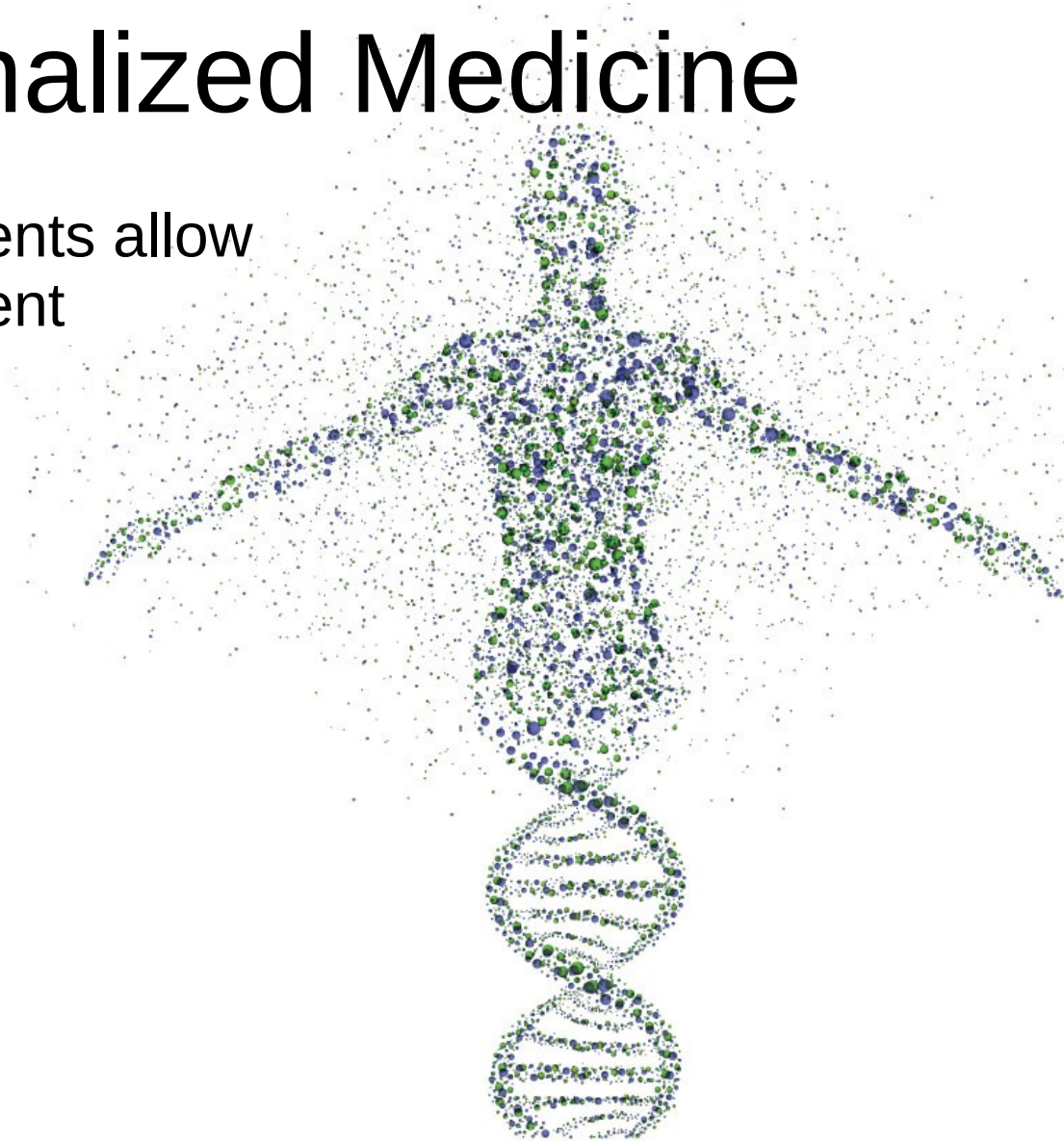


Donnelly Centre
Cellular & Biomolecular Research
UNIVERSITY OF TORONTO

UNIVERSITY OF TORONTO

Personalized Medicine

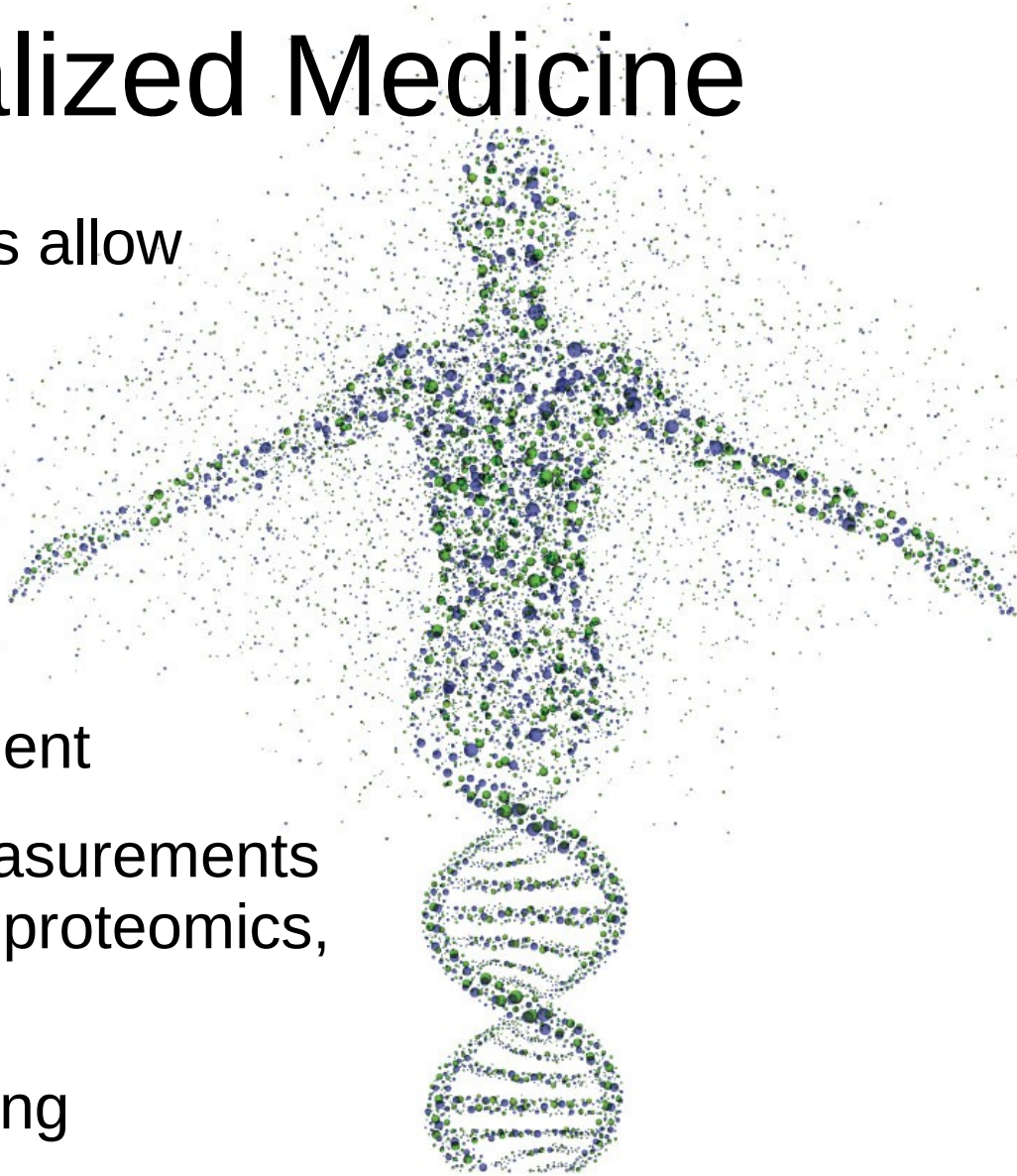
- Molecular measurements allow tailored risk assessment and therapy
- Billions of data points for each patient



Personalized Medicine

- Molecular measurements allow tailored risk assessment and therapy
- Billions of data points for each patient
- Personalized genomes allow *static* risk assessment
- Continuous *dynamic* measurements through high-throughput proteomics, metabolomics etc.

  ⇨ functional understanding



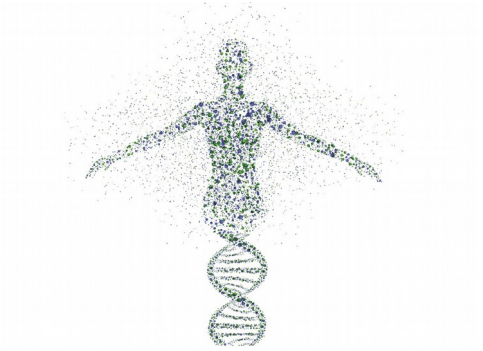
Background

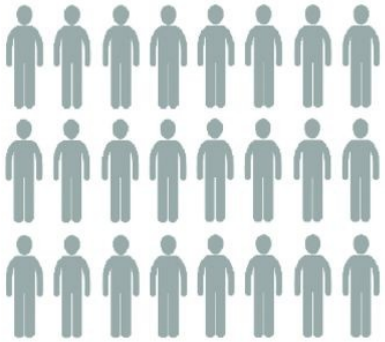
Multi-omics data

Multi-omics data

Multi-modal data becomes more common

- Same origin (same tissue / cell line / micro-organism):
  - Functional analysis
- Same patient (different origin: blood, urine, etc)
  - Correlative analysis
- Partially overlapping cohorts (missing data):
  - Imputation using shared modality

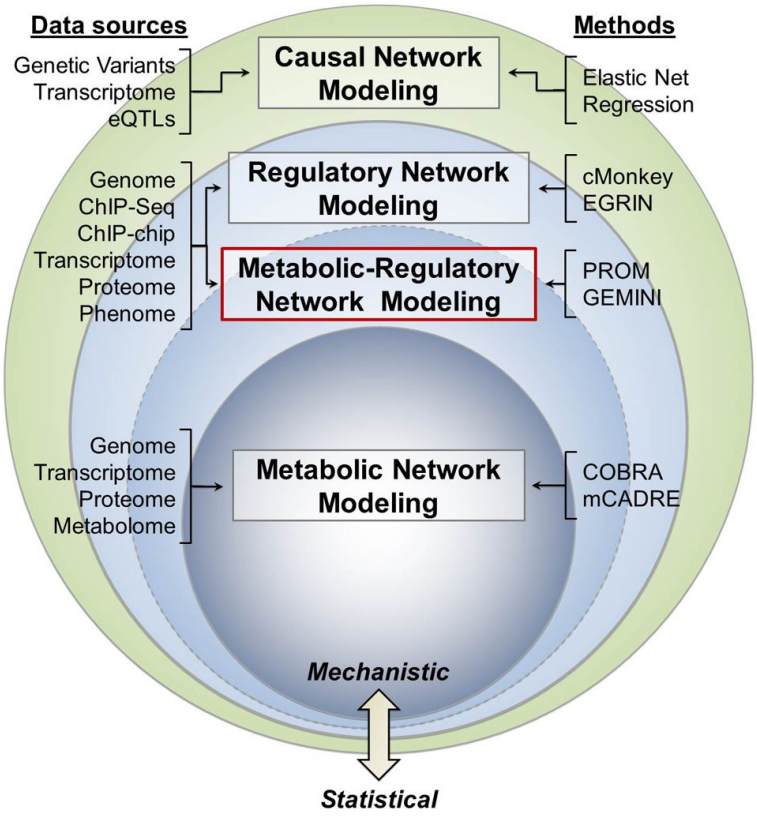




https://www.labdepotinc.com/c-439-cell-culture-flasks.php
http://www.haematologica.org/content/100/1/3.figures-only

Multi-omics data

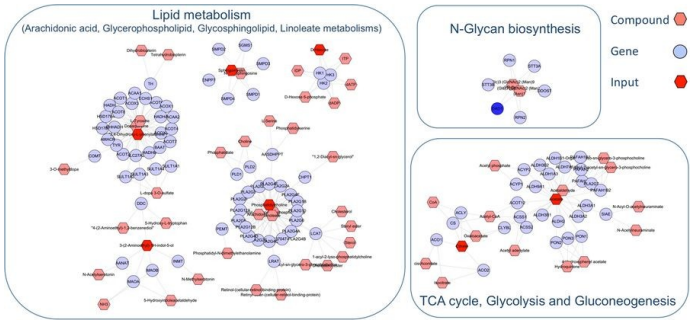
- Same origin (same tissue / cell line / micro-organism):
  - Functional analysis
  - Network based integration (transcriptomic, proteomics, metabolomic)
  - Enrichment analysis (network, GO, ...)

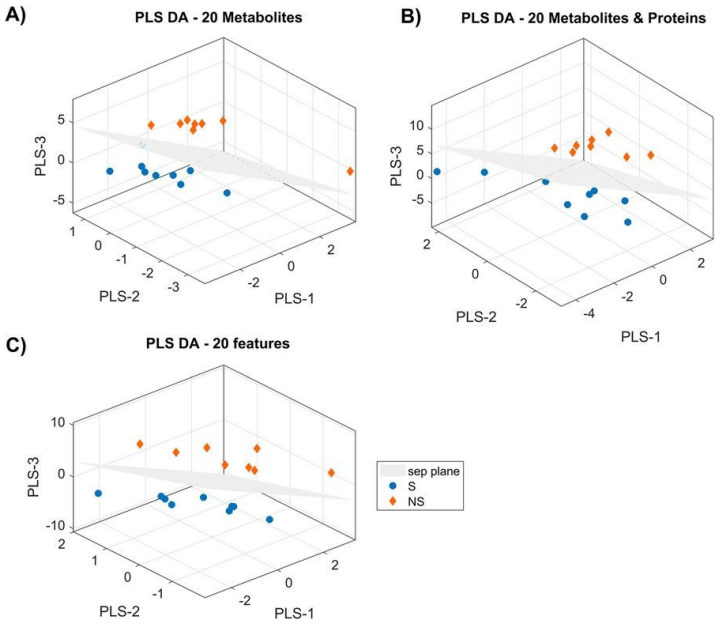


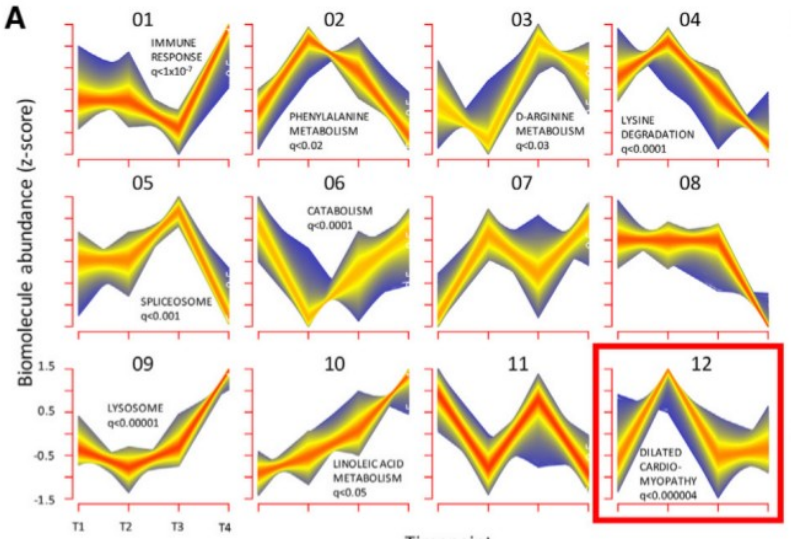
https://hood-price.systemsbiology.org/research/integrated-network-modeling/

Multi-omics data

- Same patient (different origin: blood, urine, etc)
  - Correlative analysis
  - Clustering analysis
  - Concurrent feature selection



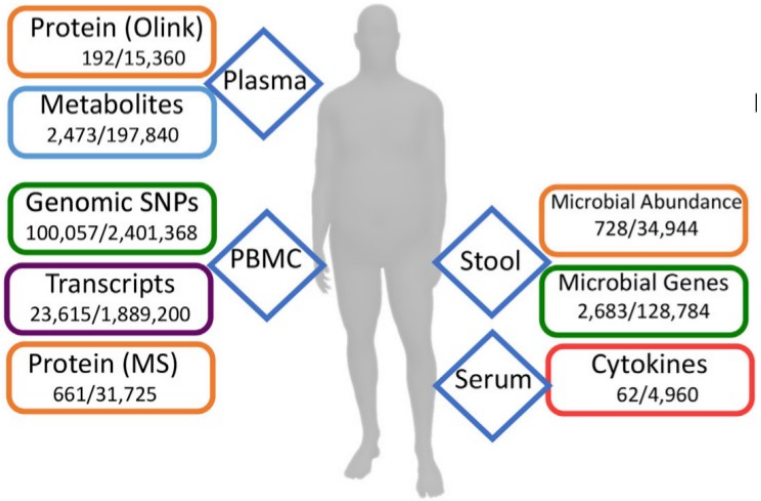




Bahado-Singh et al. Scientific Reports 2017.
Piening et al. Cell Systems 2018.
Cambiaghi et al. Scientific Reports. 2018

Multi-omics data

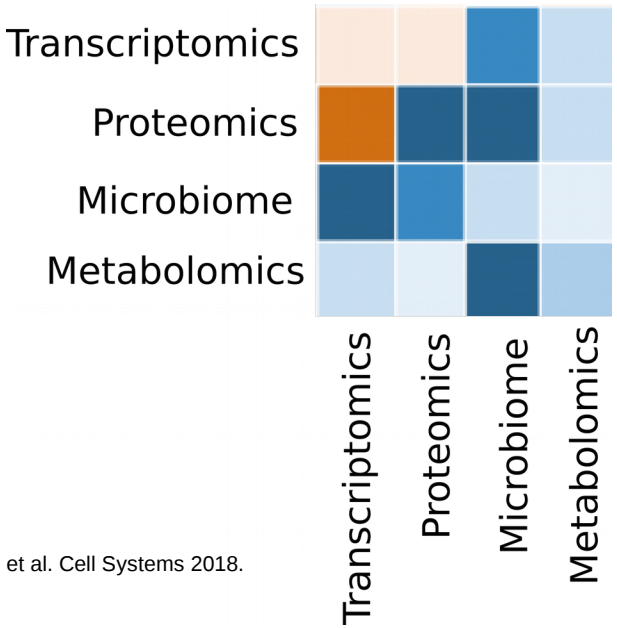
- Multi-modal data
  - Different sampling rates
  - Different number of data points (unequal contribution)
  - Highly heterogeneous data (imaging, wearables, molecular)
  - Examples:
    - iPOP (diabetes): microbiome, proteome, genome, metabolome, proteome
    - Alzheimer's Disease Neuroimaging Initiative (omics, neuro-imaging, longitudinal clinical data)
    - Parkinson's Progression Markers Initiative (omics, neuro-imaging, longitudinal clinical data)
    - All-of-us cohort (omics, behavioral, EMRs, environmental data)

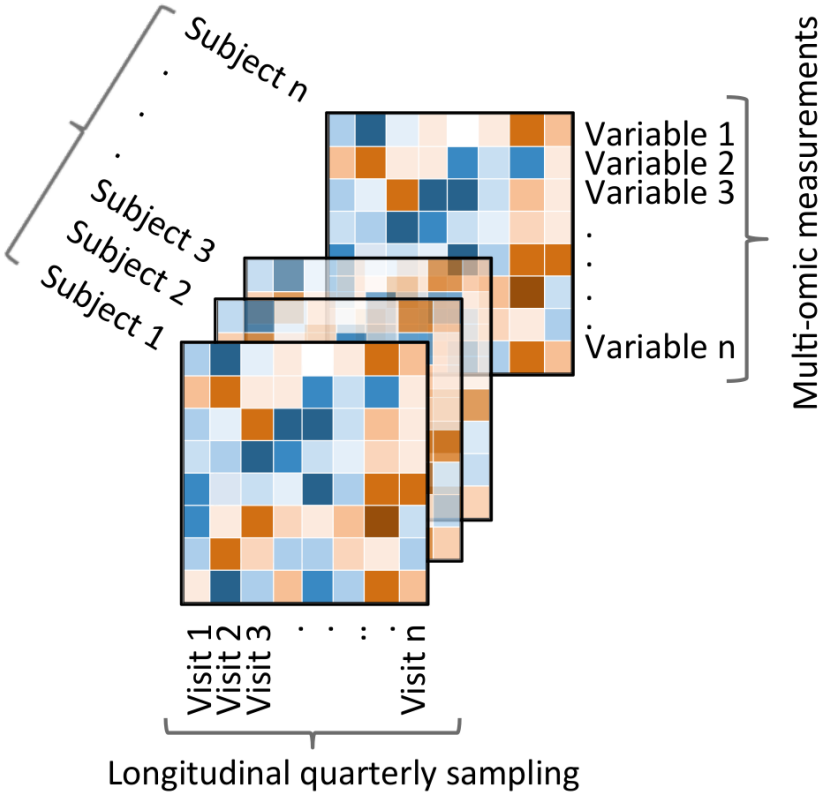


Piening et al. Cell Systems 2018.

Multi-omics data

- Why is this difficult?
  - Detection of interactions has low power (GWAS)
  - Longitudinal data is multi-dimensional
  - Interactions are N^2





Piening et al. Cell Systems 2018.

Application

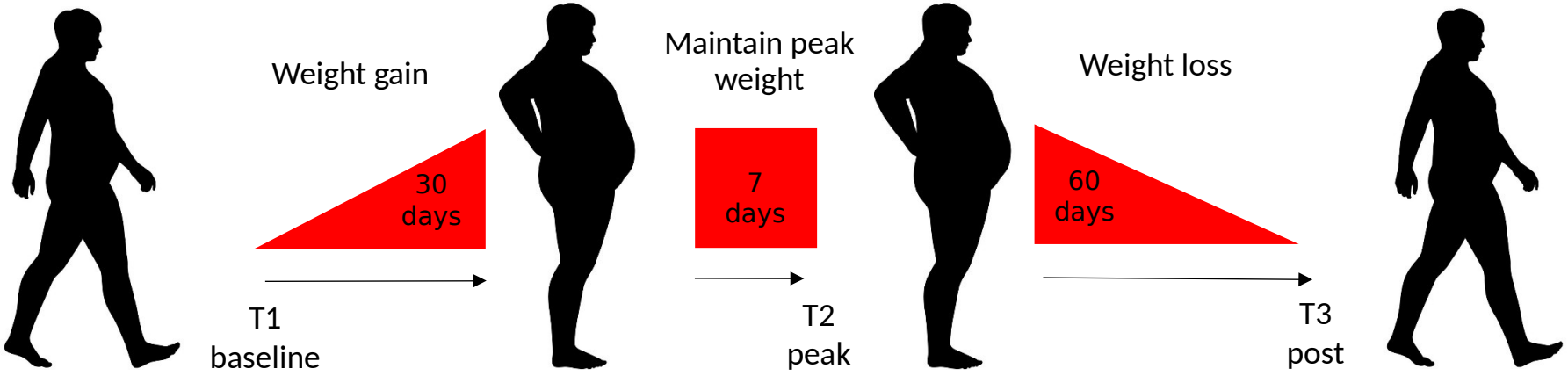
Blood plasma profiling in personalized medicine

Personalized profiling

- iPOP study with > 100 individuals
  - Generally healthy individuals with pre-diabetes
  - Profiled over multiple years (including perturbations)
- Proteomics and Metabolomics profiling with LC-MS/MS
  - Further omics: Transcriptome, Microbiome, Cytokines ...

Personalized profiling

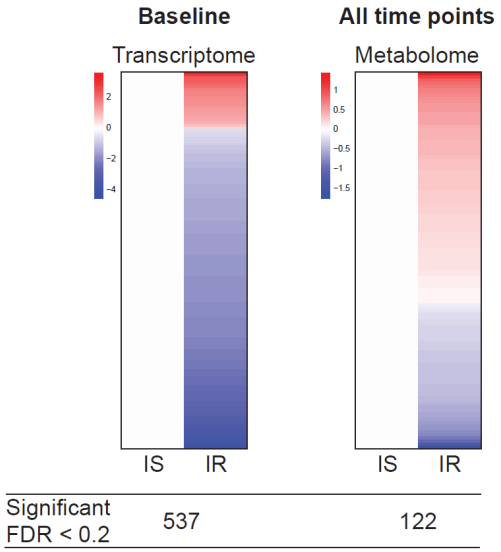
- iPOP study with > 100 individuals
  - Generally healthy individuals with pre-diabetes
  - Profiled over multiple years (including perturbations)
- Proteomics and Metabolomics profiling with LC-MS/MS
  - Further omics: Transcriptome, Microbiome, Cytokines ...
- Weight gain perturbation of IR and IS subjects (n=23)



Piening, Zhou, Contrepois, Rost *et al.* (Cell Systems, 2018)

Longitudinal profiling of Diabetes

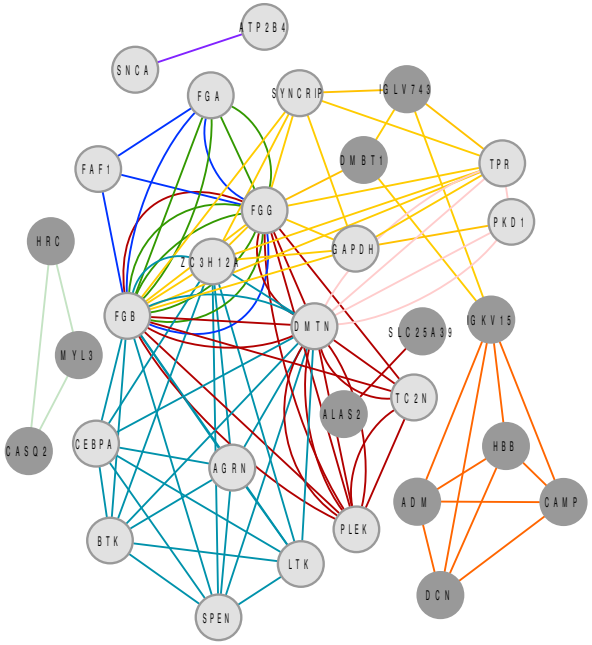
- Baseline comparison between IR and IS subjects

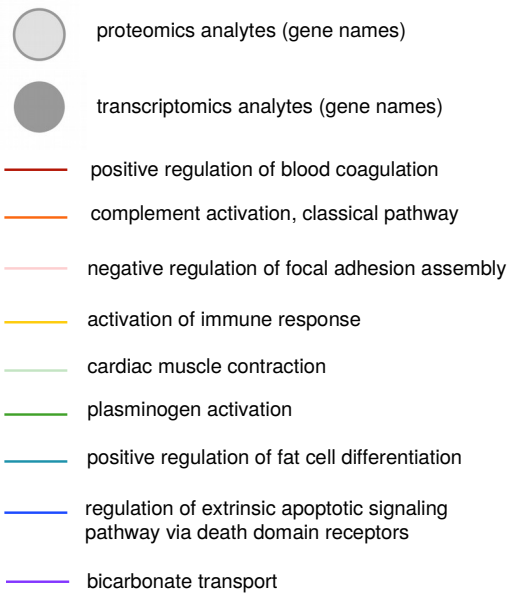


Significant FDR < 0.2	537	122

Metabolic pathways

Pathway	Database	FDR
Arginine and proline metabolism	HMDB	0.000
Glycine, serine and threonine metabolism	HMDB	0.001
Valine, leucine and isoleucine biosynthesis	HMDB	0.010
Cyanoamino acid metabolism	HMDB	0.022
Glycerophospholipid metabolism	HMDB	0.030

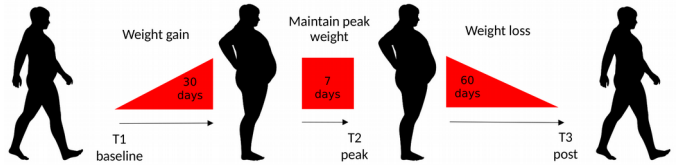


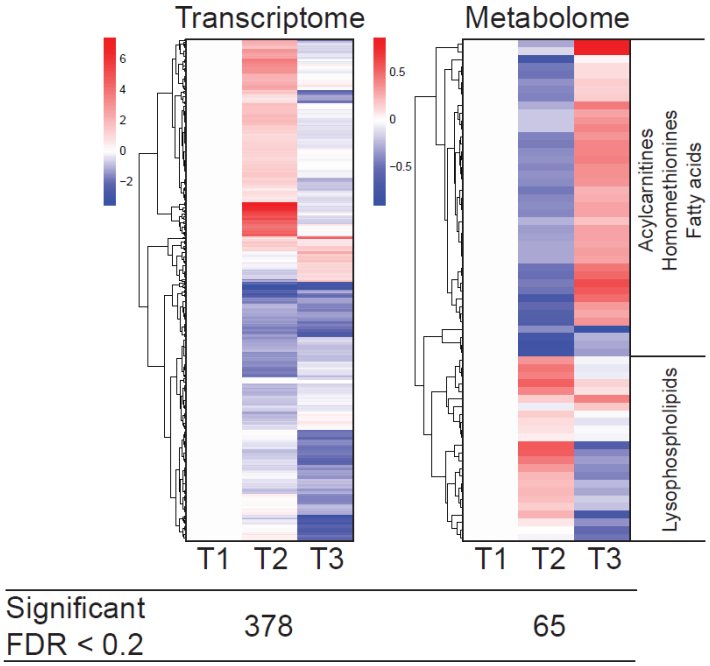


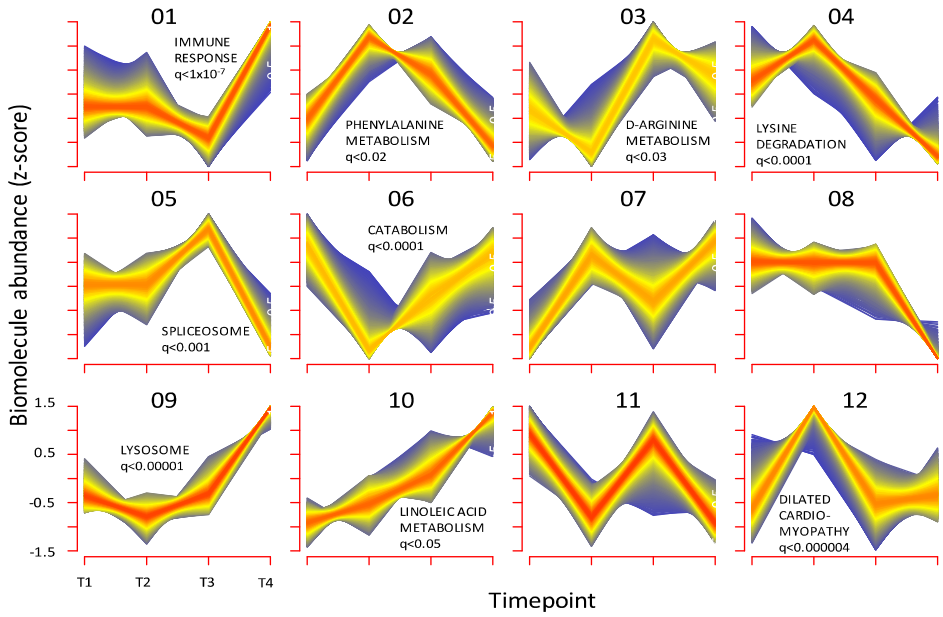
Piening, Zhou, Contrepois, Rost *et al.* (Cell Systems, 2018)

Longitudinal profiling of Diabetes

- Timepoint comparison







⇨ data analysis in temporal dimension

Piening, Zhou, Contrepois, Rost *et al.* (Cell Systems, 2018)

Outlook

Conclusions

Conclusions & Summary

- **Multi-omics analysis:** There are multiple levels of multi-omics data

- **Multi-omics longitudinal data:** Combining multiple omics over time provides insight into inter-subject variation
- **Diabetes profiling:** Analysis of baseline time-point revealed consistent differences between IR and IS subjects (AA metabolism and inflammation)
- **Personalized analysis:** Comparison to baseline timepoint provides increased statistical power

Acknowledgements

Diabetes study: Brian Piening, Wenyu Zhou, Kevin Contrepois, Gucci Gu, Tejaswini Mishra, Jessilyn Dunn, Reza Sailani, Shannon Rego, Jessica Sibal, Varsha Rao, Denis Salins, Andrew Lipchick, Liang Liang, Can Cenik, Anil Narasimha, Rohith Srivas, Christine Yeh

EMBO

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Mike Snyder

Lipidomics Integration in Multi-omics Studies: Prospects and Challenges

Jeremy P. Koelmel, PhD
jeremykoelmel@gmail.com
Adjunct Research Scientist, University of Florida

06/06/2018

Summary of Prospects

Prospects of Lipidomics:

1000+ bioactive lipids to date (functional consequences, **drug targets**)

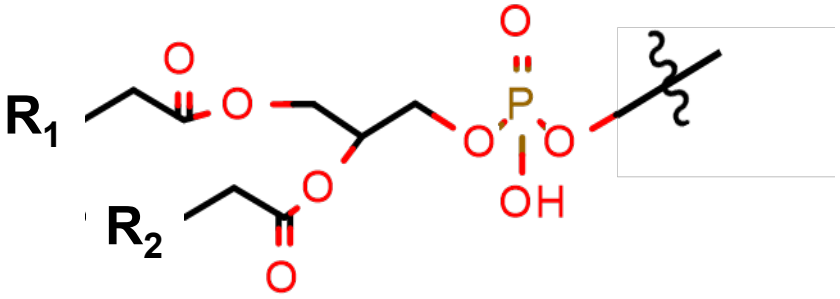
Involved in numerous disease states (**likely biomarkers**)

Same lipids across certain species (**translation to human models**)

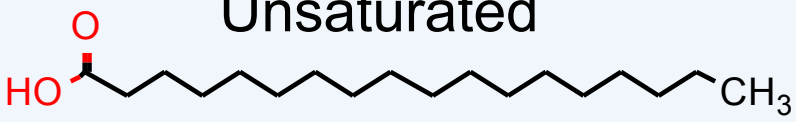
Ubiquitous, highly concentrated, high ionization efficiencies

The Actual and Possible Number of Lipids is Immense: We Cannot Map the Entire Lipidome

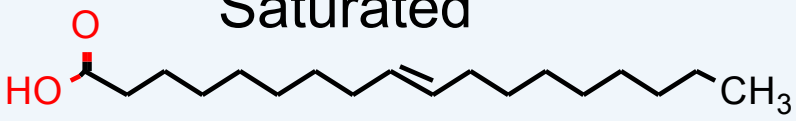
"Every week there is a report of a novel lipid being found in some exotic organism. Perhaps more surprising is how often new lipid structures are revealed in human tissues…"
Bill Christie



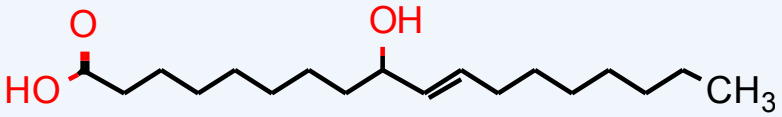
Unsaturated



Saturated

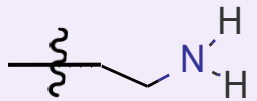


Oxidized

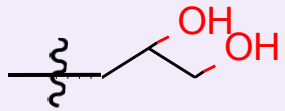


~7000 fatty acyl lipids in LipidMaps

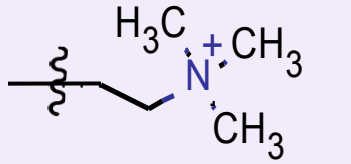
acid



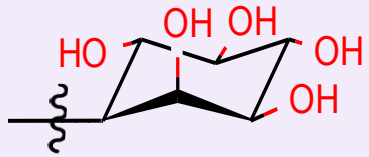
ethanolamine



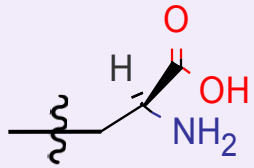
glycerol



choline



inositol

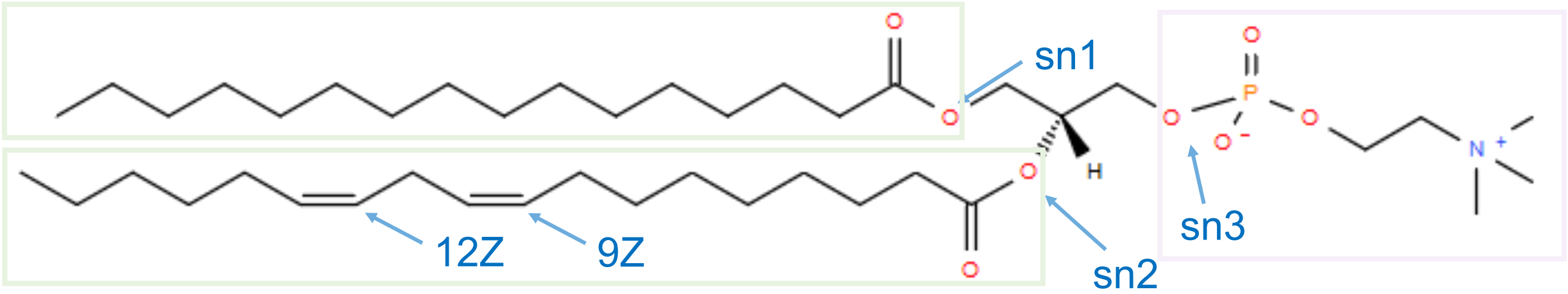


serine

...

~10000 Glycerophospholipids in LipidMaps

Capturing Lipid Diversity for Pathways Analysis



Stephen Blanksby:
“Every time you think you have one lipid you actually have 3 or 4”

Chemical Identifier: Over 143 formats!

SMILES, SYBL… KEG

SYBL…

SYBL…

SYBL, InChI, SMILES…

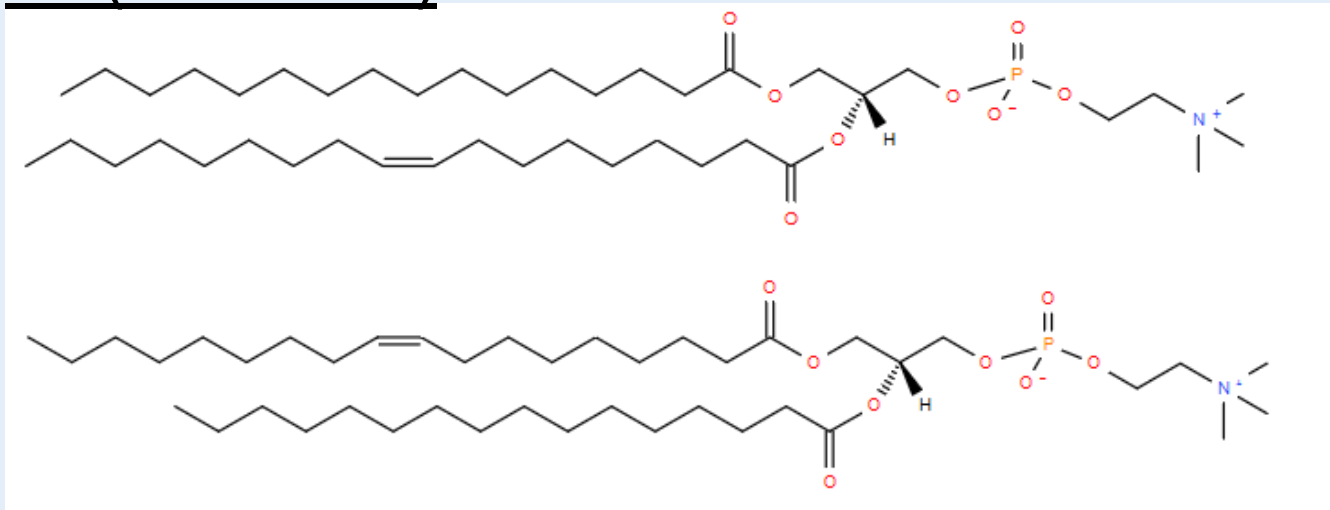
SYBL, InChI, SMILES… KEG

Subtle Structural Changes and Biological Importance

Double bond position, cis versus trans double bond, fatty acid position, fatty acid chain length

Membrane fluidity

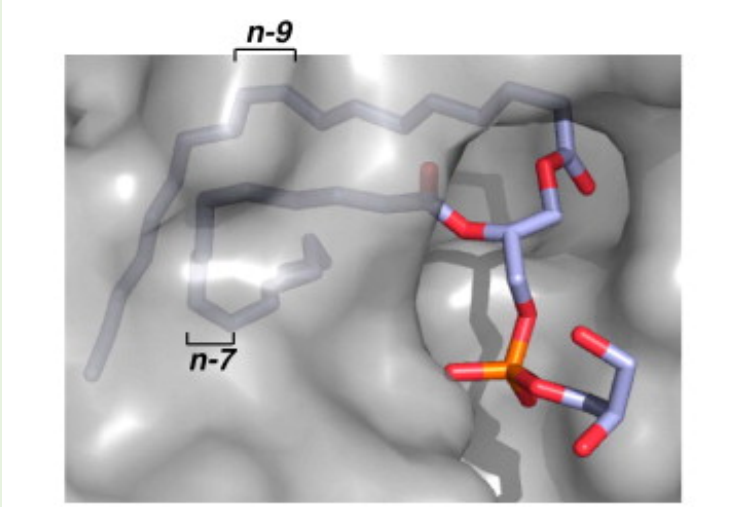
PC(16:0/18:1)



PC(18:1/16:0)

Protein Receptor Binding

nuclear receptor NR5A - development, homeostasis, & metabolism



Brown et al. 2011:
Analysis of unsaturated lipids by ozone-induced dissociation

Summary of Prospects and Challenges

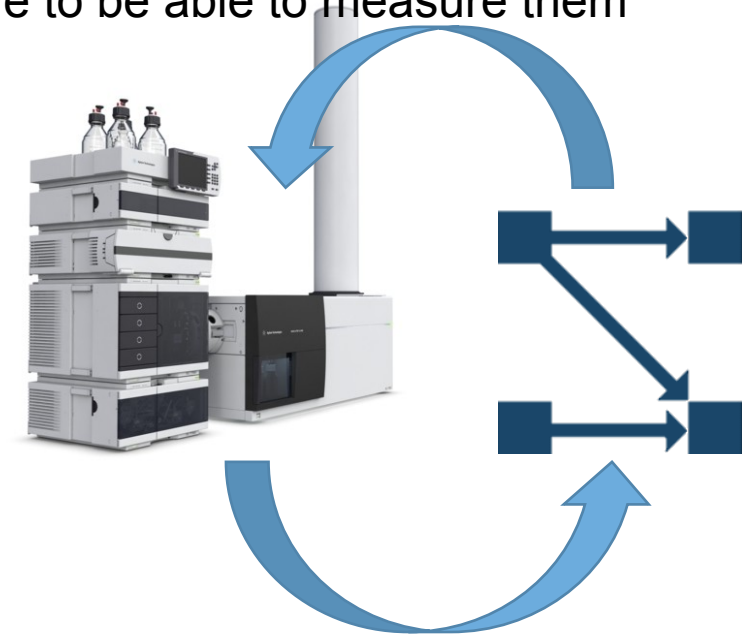
Prospects of Lipidomics:

1000+ bioactive lipids to date (functional consequences, **drug targets**)
Involved in numerous disease states (**likely biomarkers**)
Same lipids across certain species (**translation to human models**)
Ubiquitous, highly concentrated, high ionization efficiencies

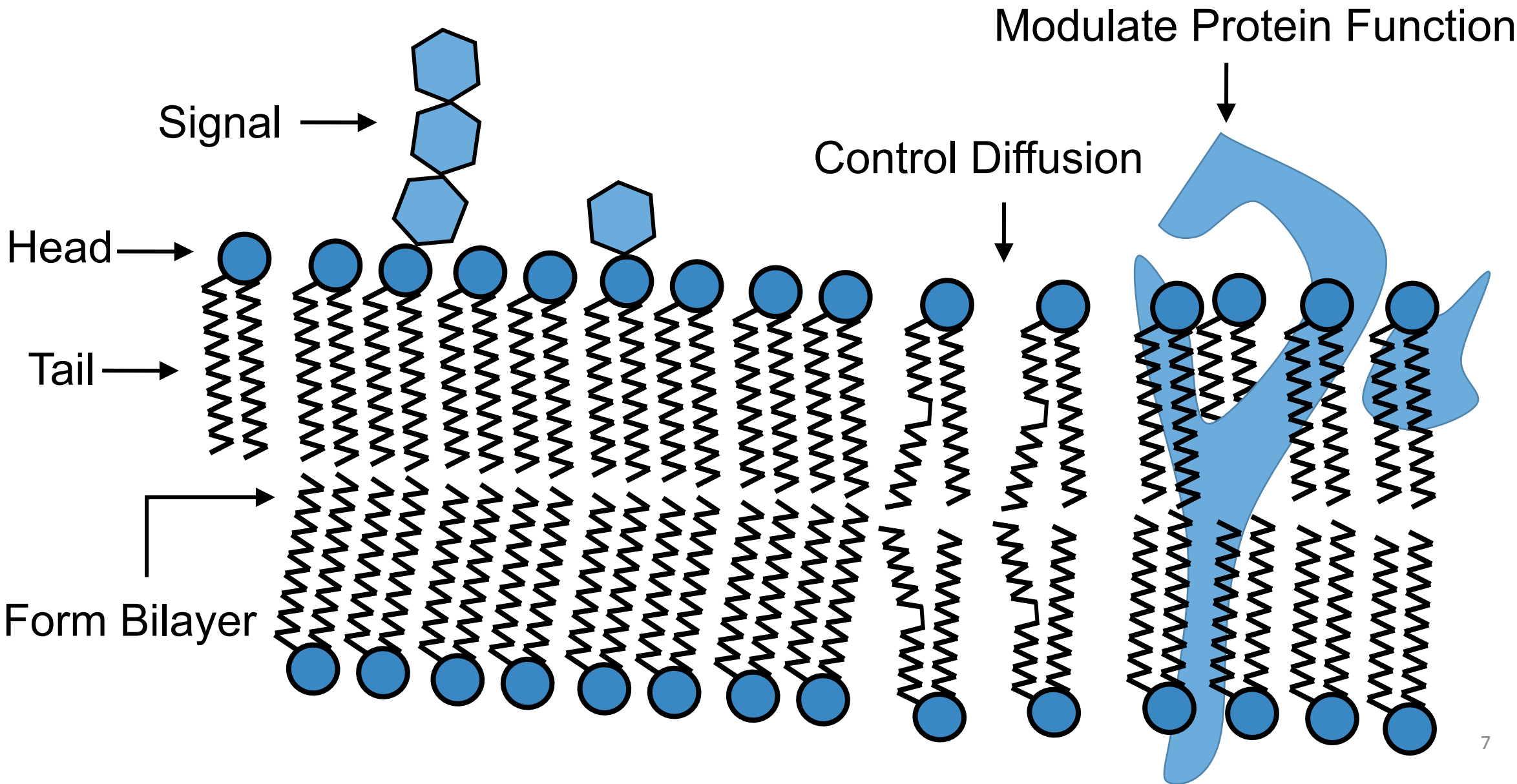
Challenges in pathway analysis:

Separating and identifying lipid isomers
Normalization/quantification of lipids (data quality)
Identifiers representing varying structural information
Limited databases for pathway analysis

Before we can know anything about lipids,
We have to be able to measure them



Lipids: Functional Diversity



Novel workflows to investigate cross-omics interactions networks

Ilaria Piazza

Picotti Group, Institute of Molecular Systems Biology, ETH Zurich

ASMS 2018, San Diego

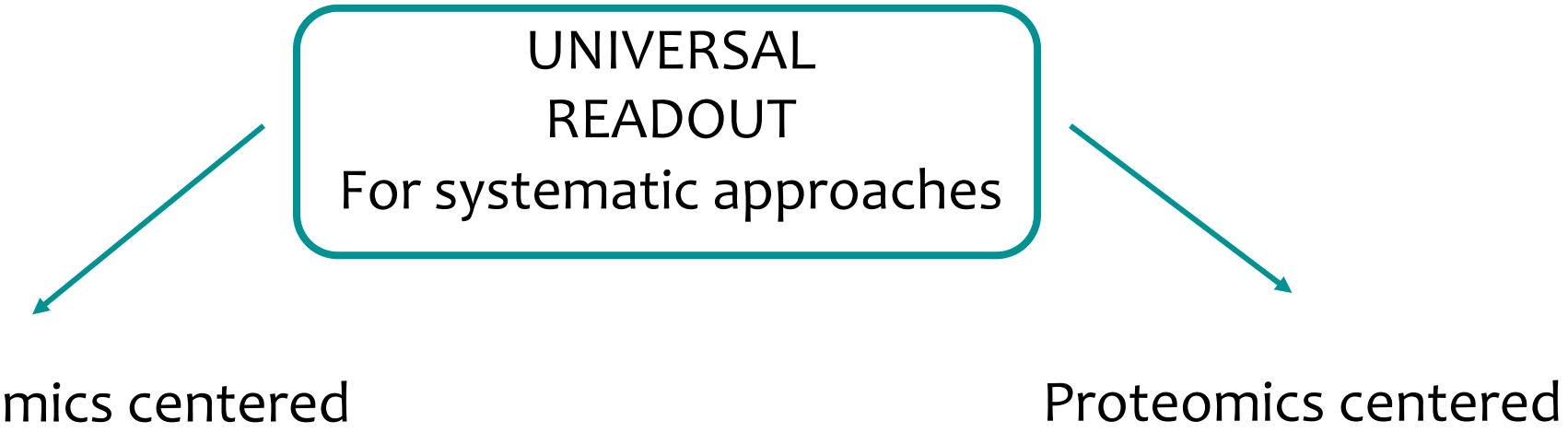
Multi-MS-omics Data integration workshop

Systematic discovery of Protein-Metabolite binding events: The current experimental bottleneck

Any compound chemistry

Proteome-wide or metabolome -wide

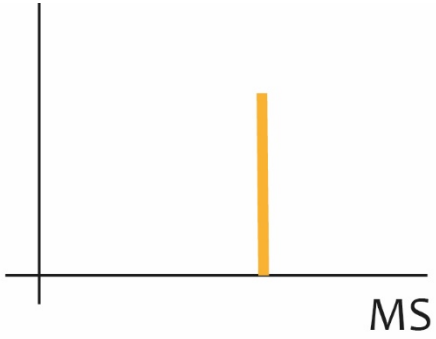
Complex cell matrices – under near physiological conditions

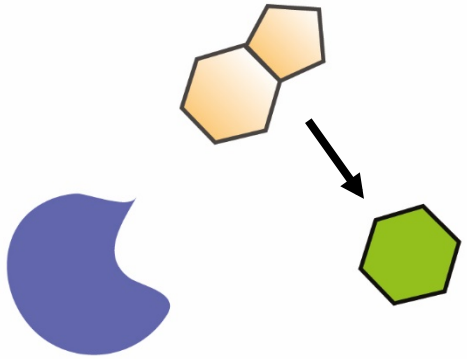


Metabolomic centered:
Measuring global metabolite binding to protein targets

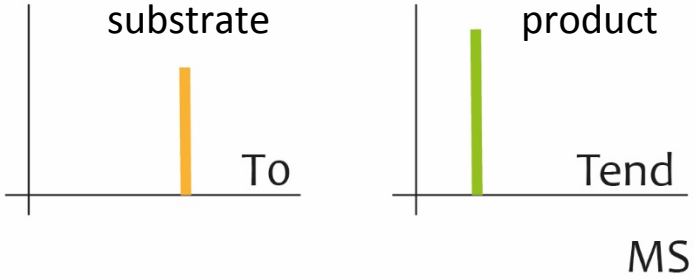
Multiplexed protein purification

Interaction





Reaction

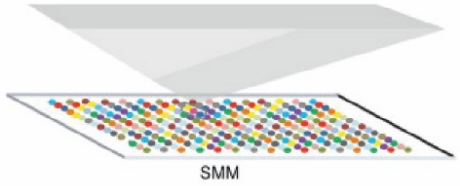


READOUT:
MS based metabolomics

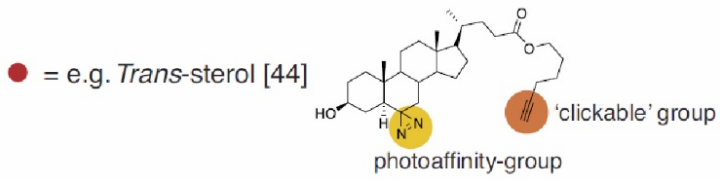
Li, et al 2010,
Sevin et. al. 2017

Proteomics centered:
Measuring global protein binding to metabolite ligands

Small molecule microarrays



Chemically functionalized small molecules



READOUT:
MS based proteomics

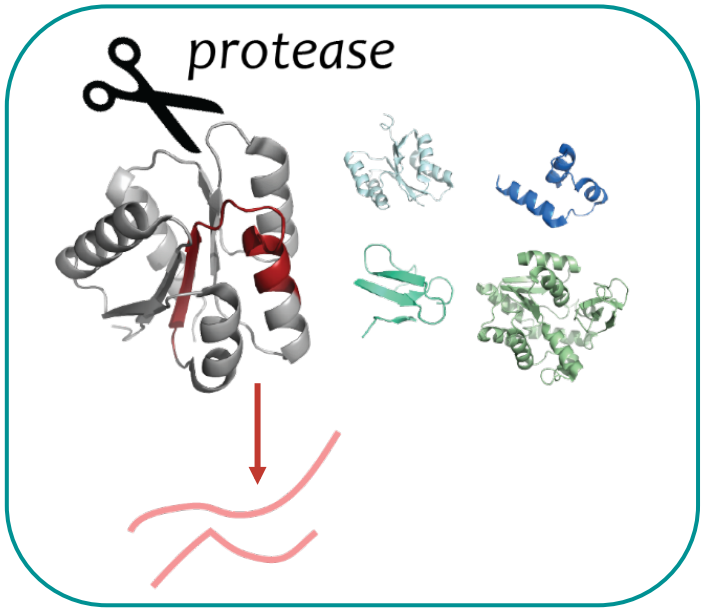
Proteome-wide measurements of Structural states

- LiP-SMap

Diether, et al 2017; Saliba, et al, 2016; Hulce et al 2013; Backus at al, 2016; Savitski et al 2014

Systematic discovery of Protein-Metabolite binding events with **structural proteomics**

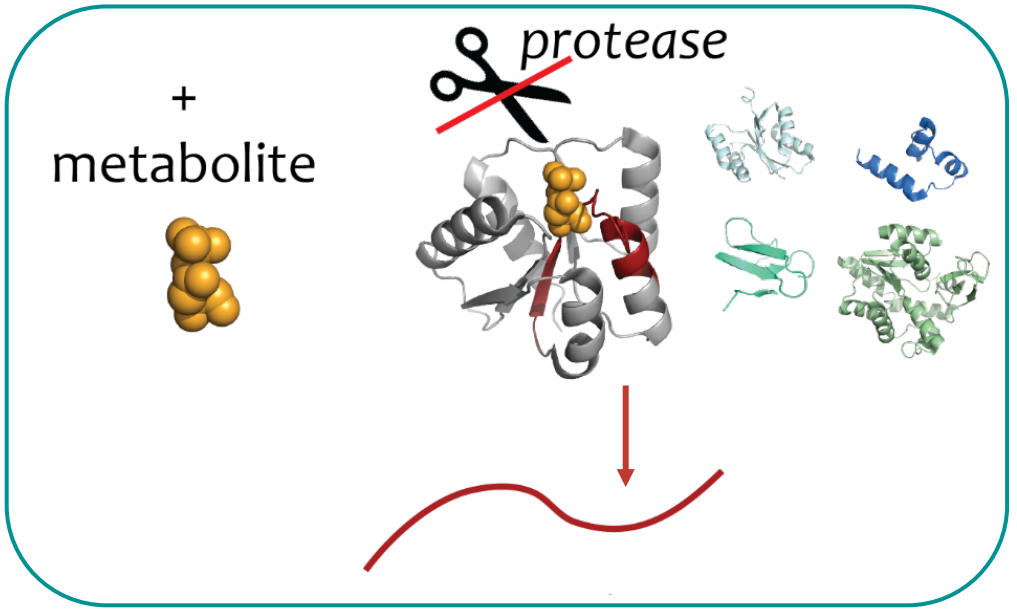
- ✓ Any compound chemistry
- ✓ Proteome-wide
- ✓ Complex cell matrices, under near physiological conditions.



Metabolite Free

READOUT
Peptide markers

LiP peptides

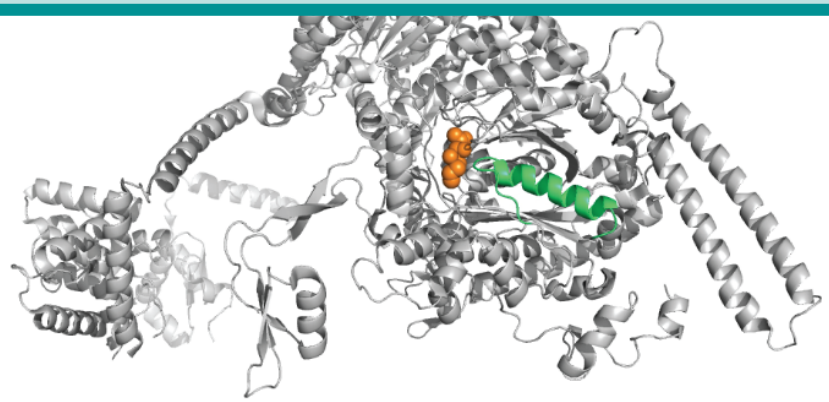


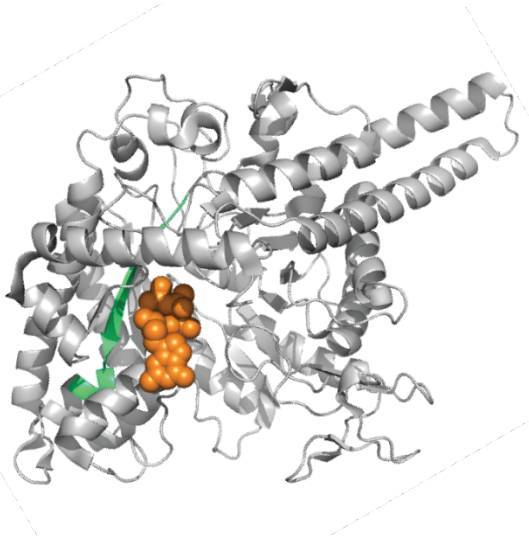
Metabolite Bound

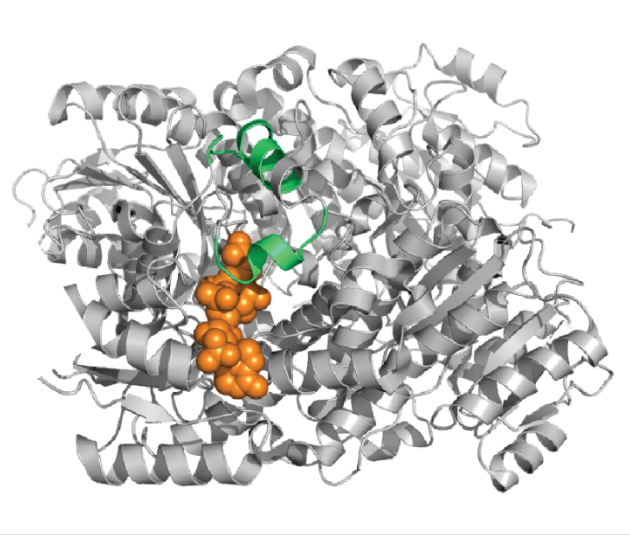
LiP-SMap method: Pinpointing metabolite binding proteins and binding sites

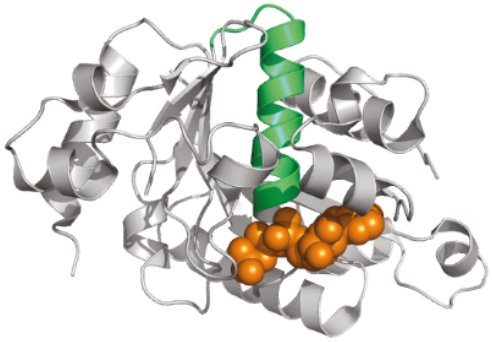
LiP Peptide markers:

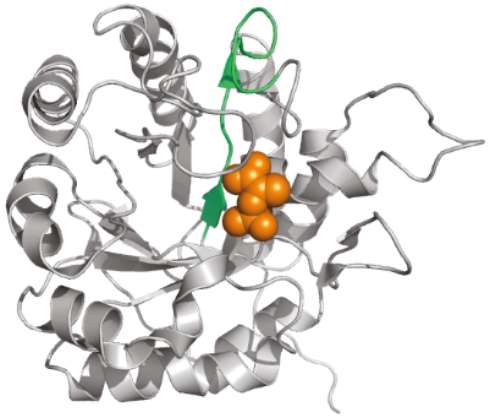
- Identify protein metabolite interactions
- Pinpoint metabolite binding sites

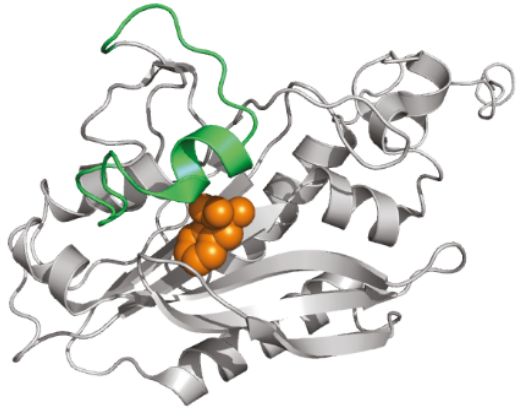




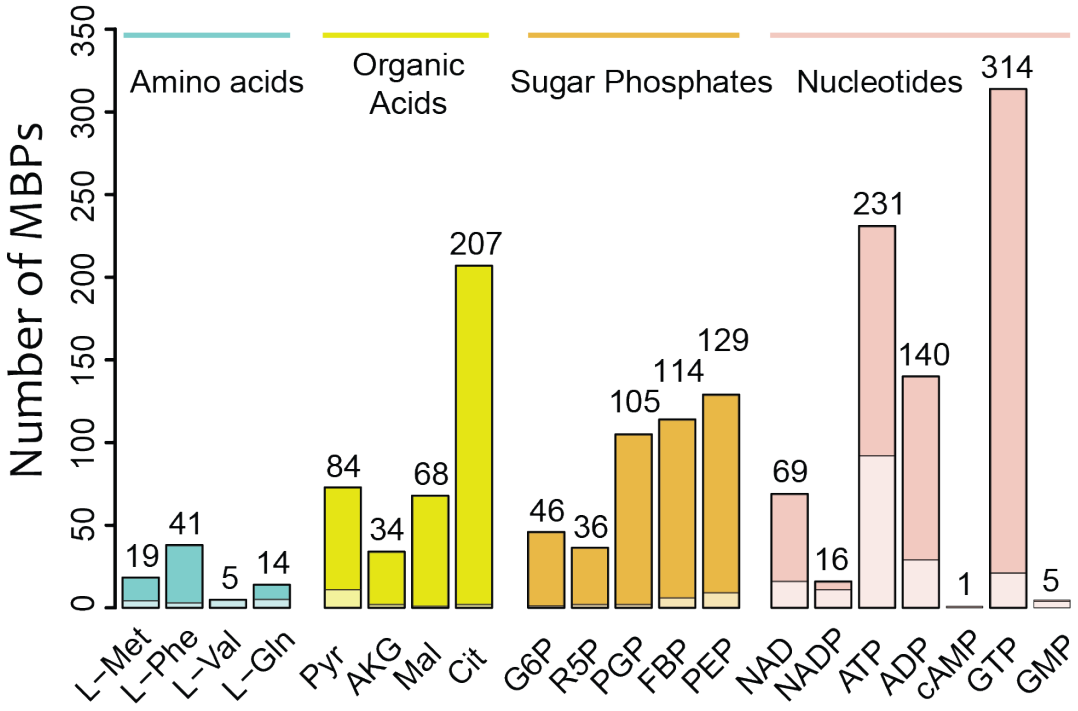




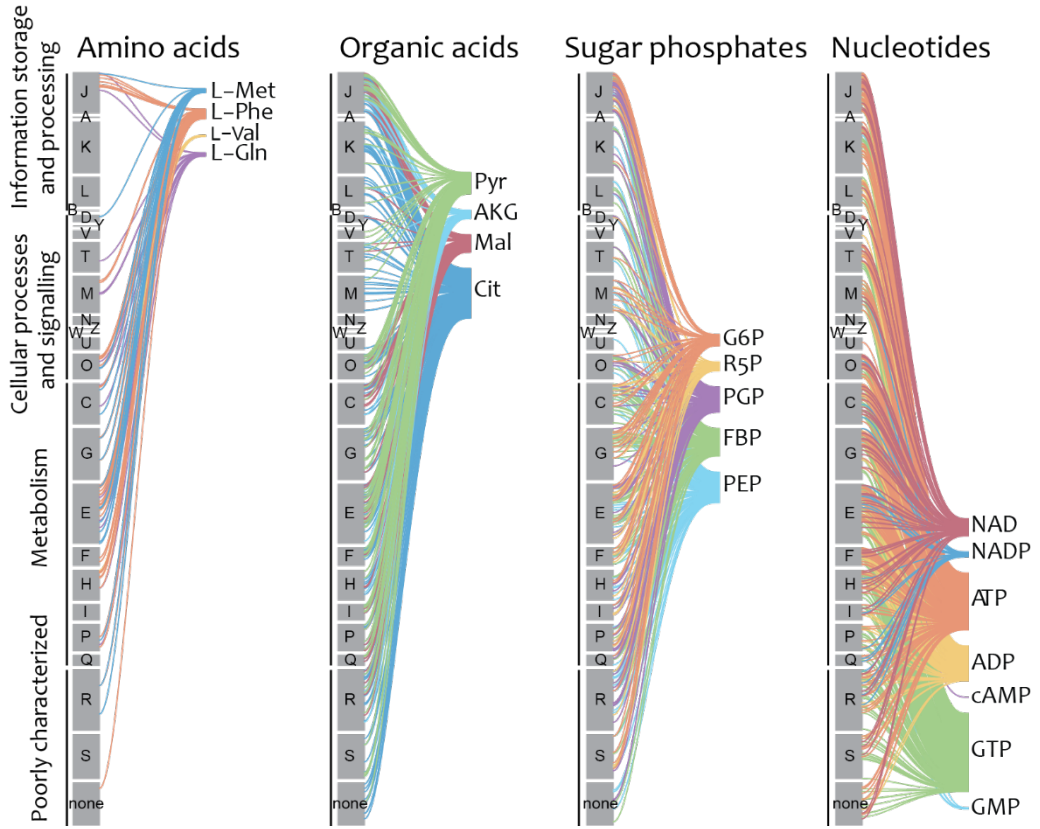


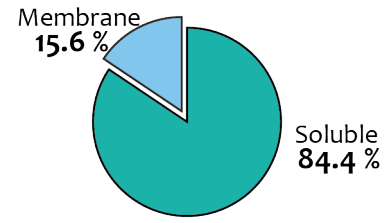


A case study for cross-omics:
The *E.coli* map of Protein-Metabolite interactions with LiP-SMap

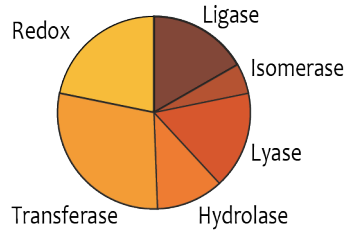


- ~ 1700 interactions (80% novel)
- 76 uncharacterized proteins
- Multiple cellular processes





Cellular compartment of Metabolite binding proteins



Catalytic activity of Metabolite binding proteins

Functional relevance of protein-metabolite interactions

Protein-metabolite interactome map

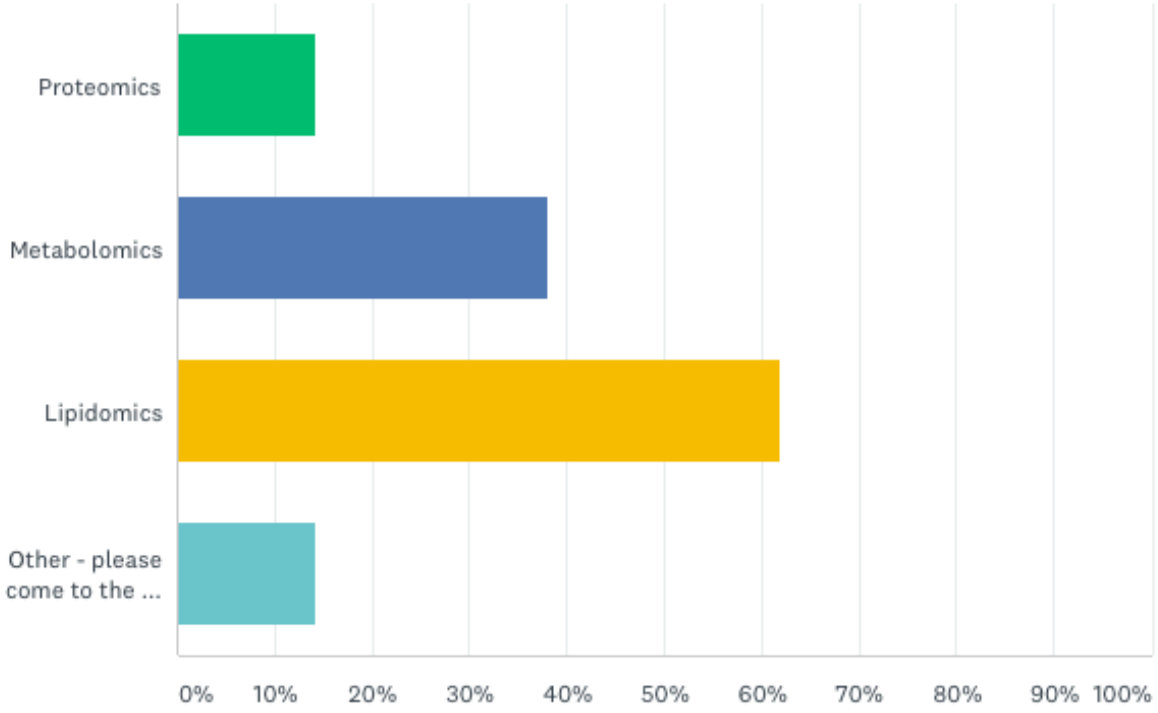
=

Matrix of binary associations

Are they physiologically relevant?

Which MS-omics technology needs the most effort for primary data analysis (moving from raw data to quantitative identifications)?

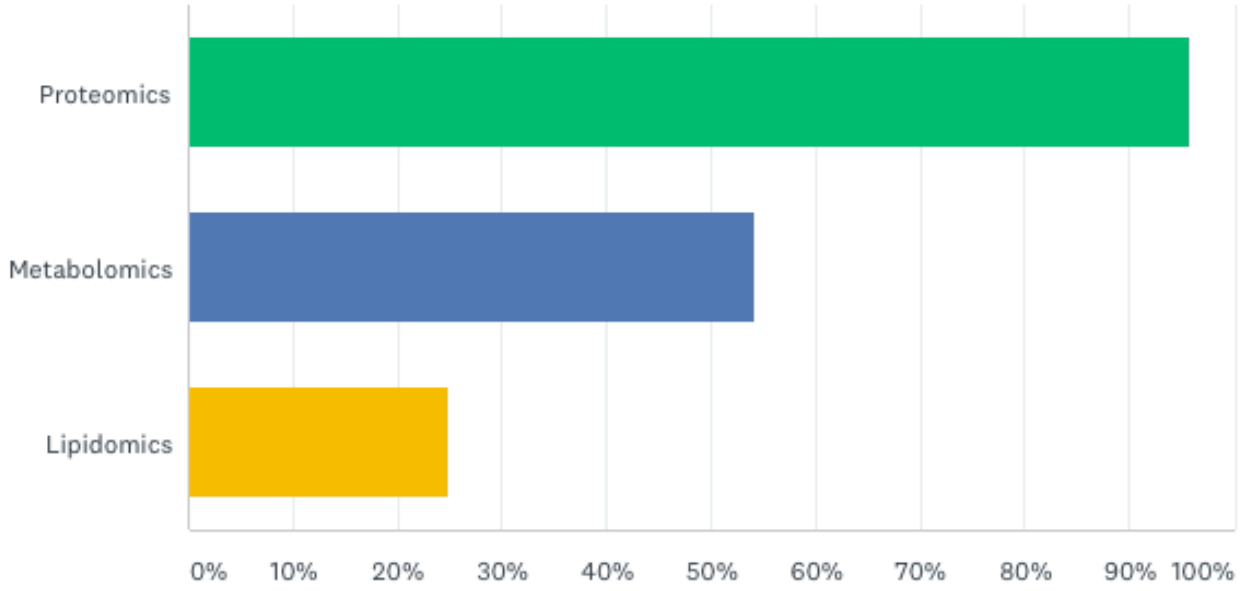
Answered: 21 Skipped: 3



ANSWER CHOICES	RESPONSES	
Proteomics	14.29%	3
Metabolomics	38.10%	8
Lipidomics	61.90%	13
Other - please come to the mic and tell us!	14.29%	3
Total Respondents: 21		

Which of the different MS-omics technologies are ready to be used in multi-MS-omics data integration studies?

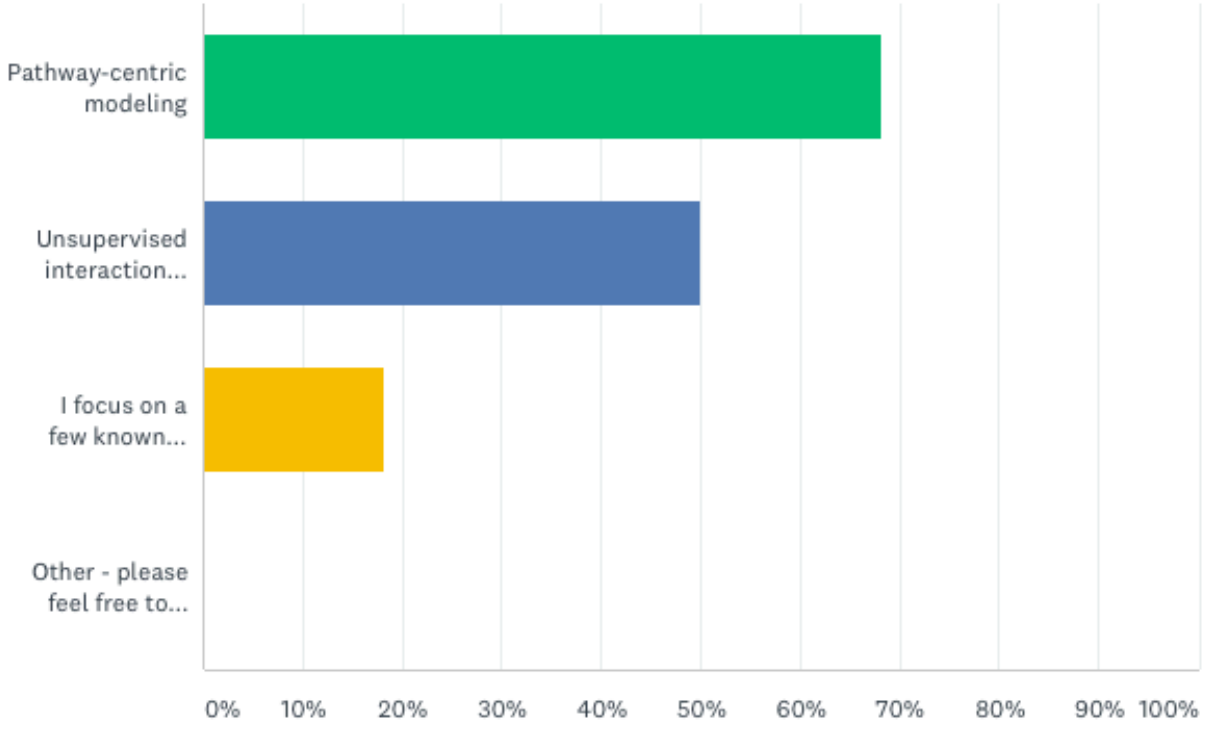
Answered: 24 Skipped: 0



ANSWER CHOICES	RESPONSES	
Proteomics	95.83%	23
Metabolomics	54.17%	13
Lipidomics	25.00%	6
Total Respondents: 24		

How do you prefer to integrate MS-omics data?

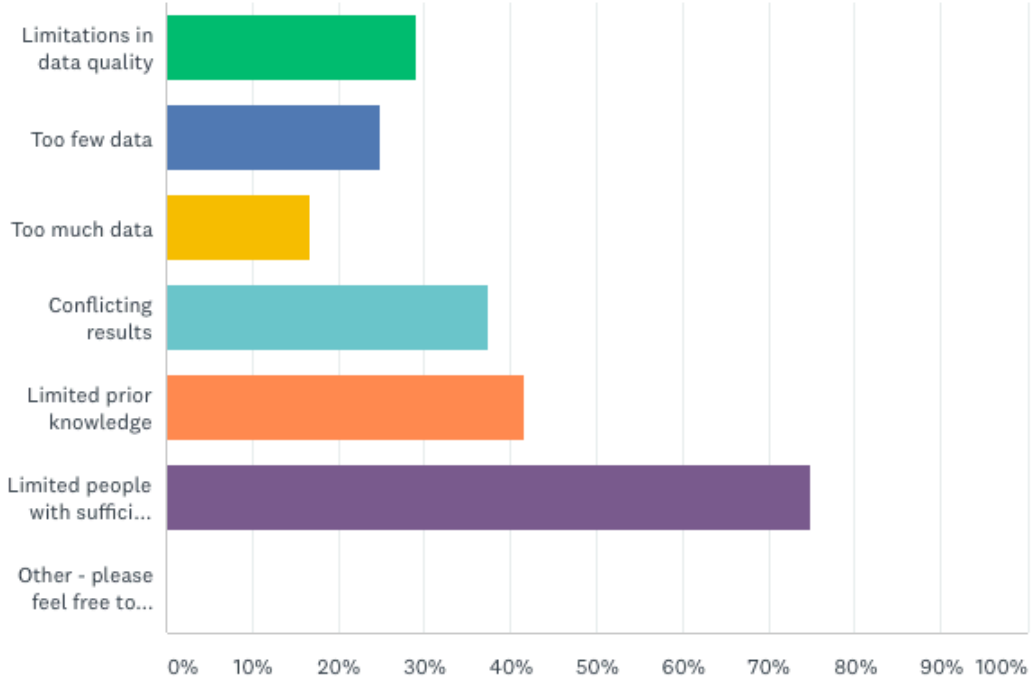
Answered: 22 Skipped: 2



ANSWER CHOICES	RESPONSES	
Pathway-centric modeling	68.18%	15
Unsupervised interaction networks	50.00%	11
I focus on a few known molecules and do integration manually	18.18%	4
Other - please feel free to share your opinion by going to the microphone!	0.00%	0
Total Respondents: 22		

What are the biggest challenges in muli-MS-omics data integration?

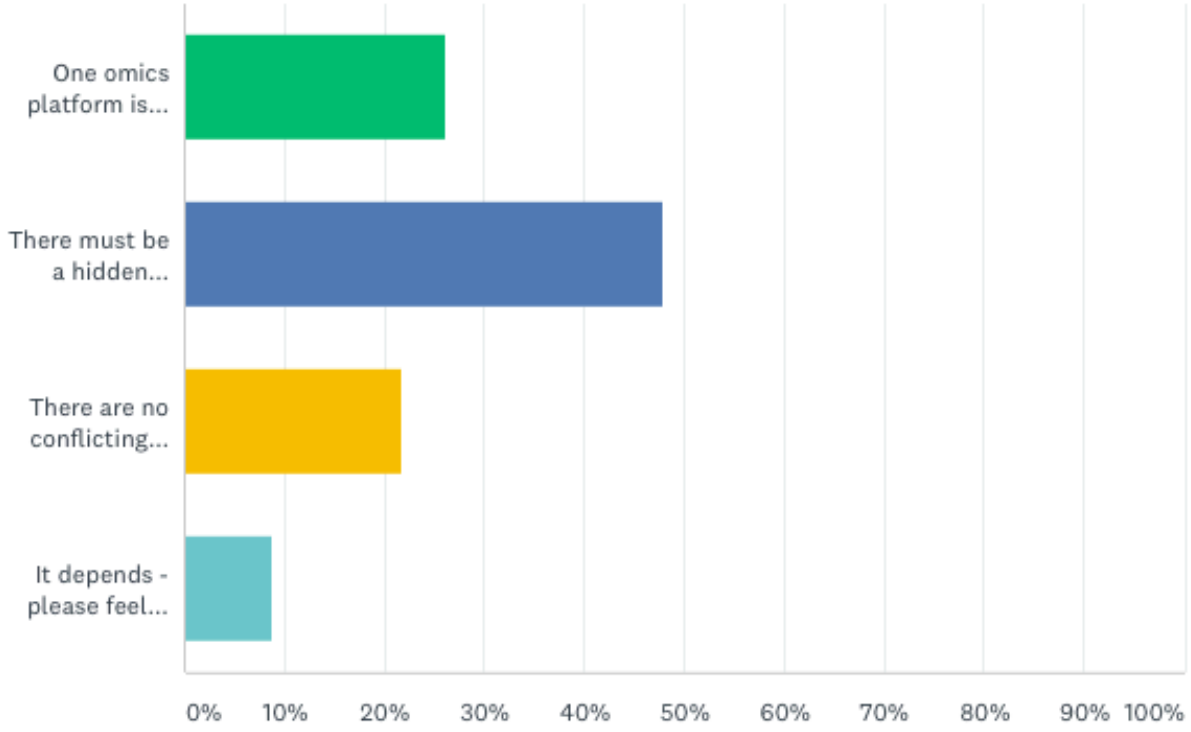
Answered: 24 Skipped: 0



ANSWER CHOICES	RESPONSES	
Limitations in data quality	29.17%	7
Too few data	25.00%	6
Too much data	16.67%	4
Conflicting results	37.50%	9
Limited prior knowledge	41.67%	10
Limited people with sufficient expertise across multiple omics technologies	75.00%	18
Other - please feel free to share your opinion by going to the microphone!	0.00%	0
Total Respondents: 24		

How to handle conflicting results?

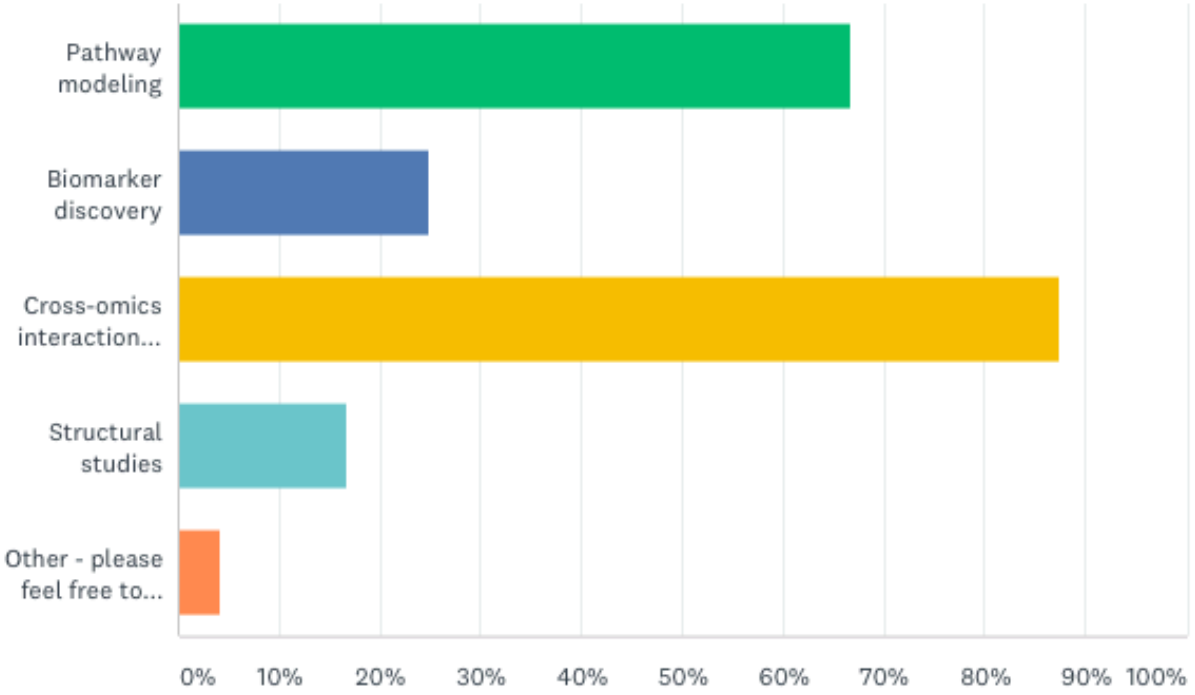
Answered: 23 Skipped: 1



ANSWER CHOICES	RESPONSES	
One omics platform is wrong - choose the one you trust most	26.09%	6
There must be a hidden regulation in between	47.83%	11
There are no conflicting results!	21.74%	5
It depends - please feel free to share your opinion by going to the microphone!	8.70%	2
Total Respondents: 23		

What are the most promising future directions of multi-MS-omics?

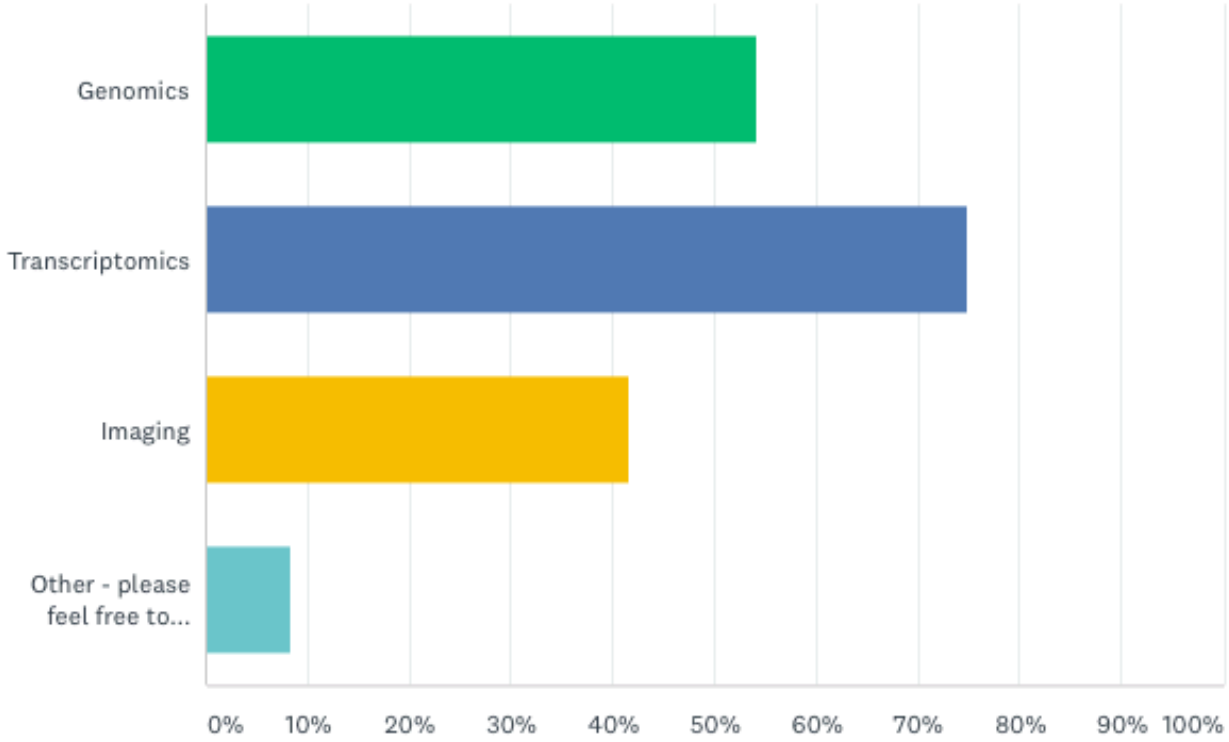
Answered: 24 Skipped: 0



ANSWER CHOICES	RESPONSES	
Pathway modeling	66.67%	16
Biomarker discovery	25.00%	6
Cross-omics interaction networks	87.50%	21
Structural studies	16.67%	4
Other - please feel free to share your opinion by going to the microphone!	4.17%	1
Total Respondents: 24		

What would be other interesting data types to integrate?

Answered: 24 Skipped: 0



ANSWER CHOICES	RESPONSES	
Genomics	54.17%	13
Transcriptomics	75.00%	18
Imaging	41.67%	10
Other - please feel free to share your opinion by going to the microphone!	8.33%	2
Total Respondents: 24		