

Omics interplay for health and disease

with focus in genetics, metagenomics, causality and longitudinal multi-omics data

Marju Orho-Melander

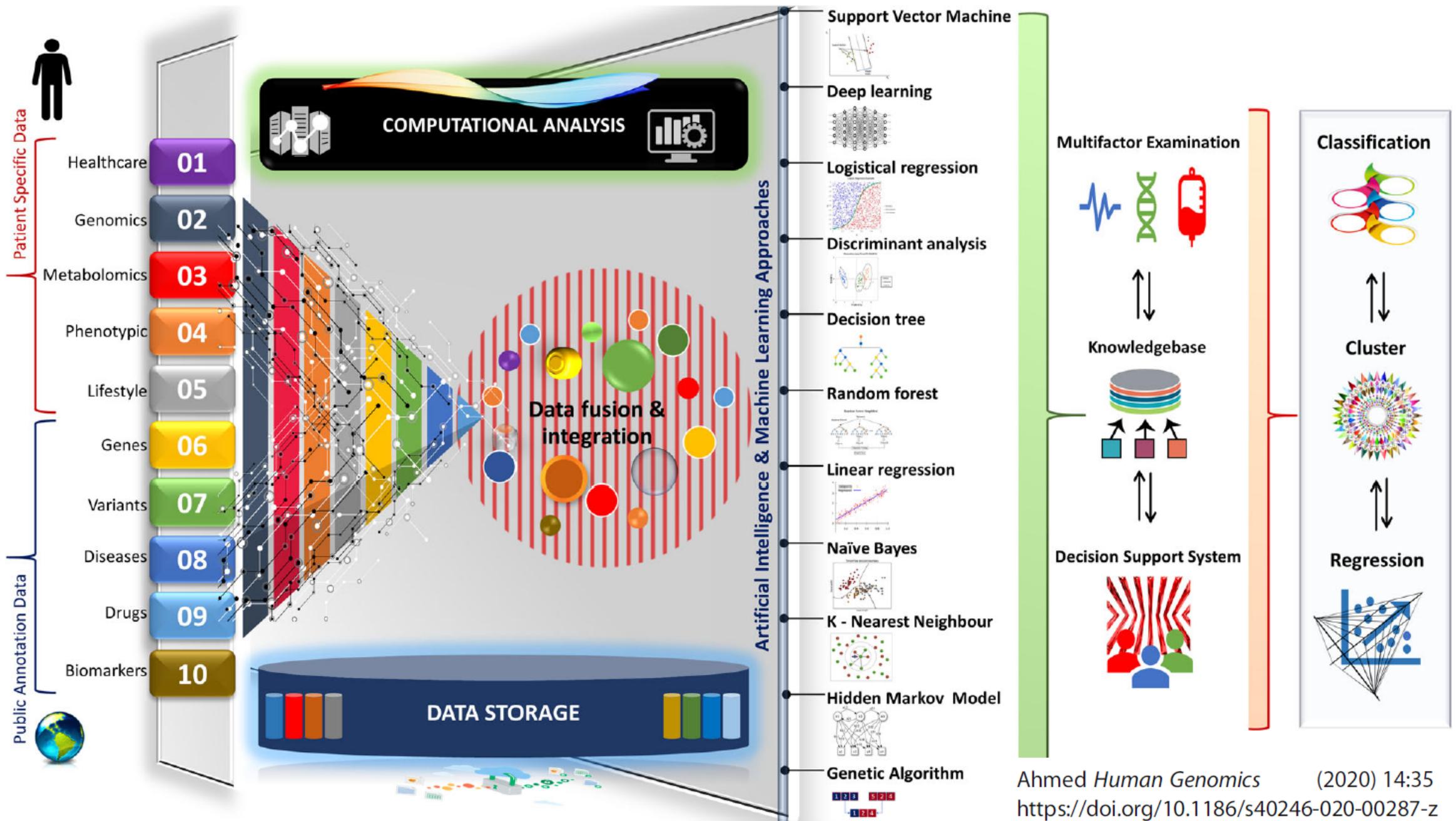
Professor of genetic epidemiology

Lund University

marju.orho-melander@med.lu.se

MY AIMS FOR THE LECTURE

- What kind of time we are living?
- Genomics, metagenomics and causality
- Some lessons from longitudinal multi-omics analysis
- Metagenomics in >12,000 Swedes
- To provide some interesting papers to read



High-Definition Medicine

Ali Torkamani,^{1,2,5,*} Kristian G. Andersen,^{1,3} Steven R. Steinhubl,^{1,4} and Eric J. Topol, MD^{1,4}

¹The Scripps Translational Science Institute, La Jolla, CA 92037, USA

²Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

³Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, CA 92037, USA

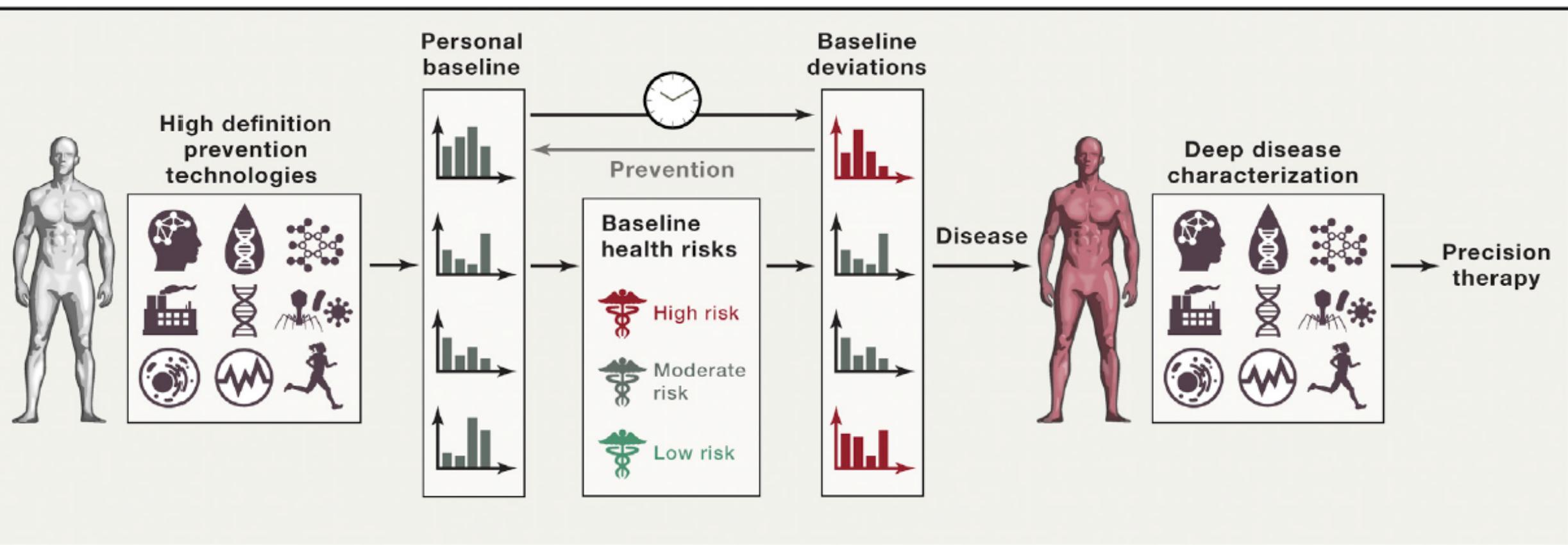
⁴Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA

⁵Lead Contact

*Correspondence: atorkama@scripps.edu

<http://dx.doi.org/10.1016/j.cell.2017.08.007>

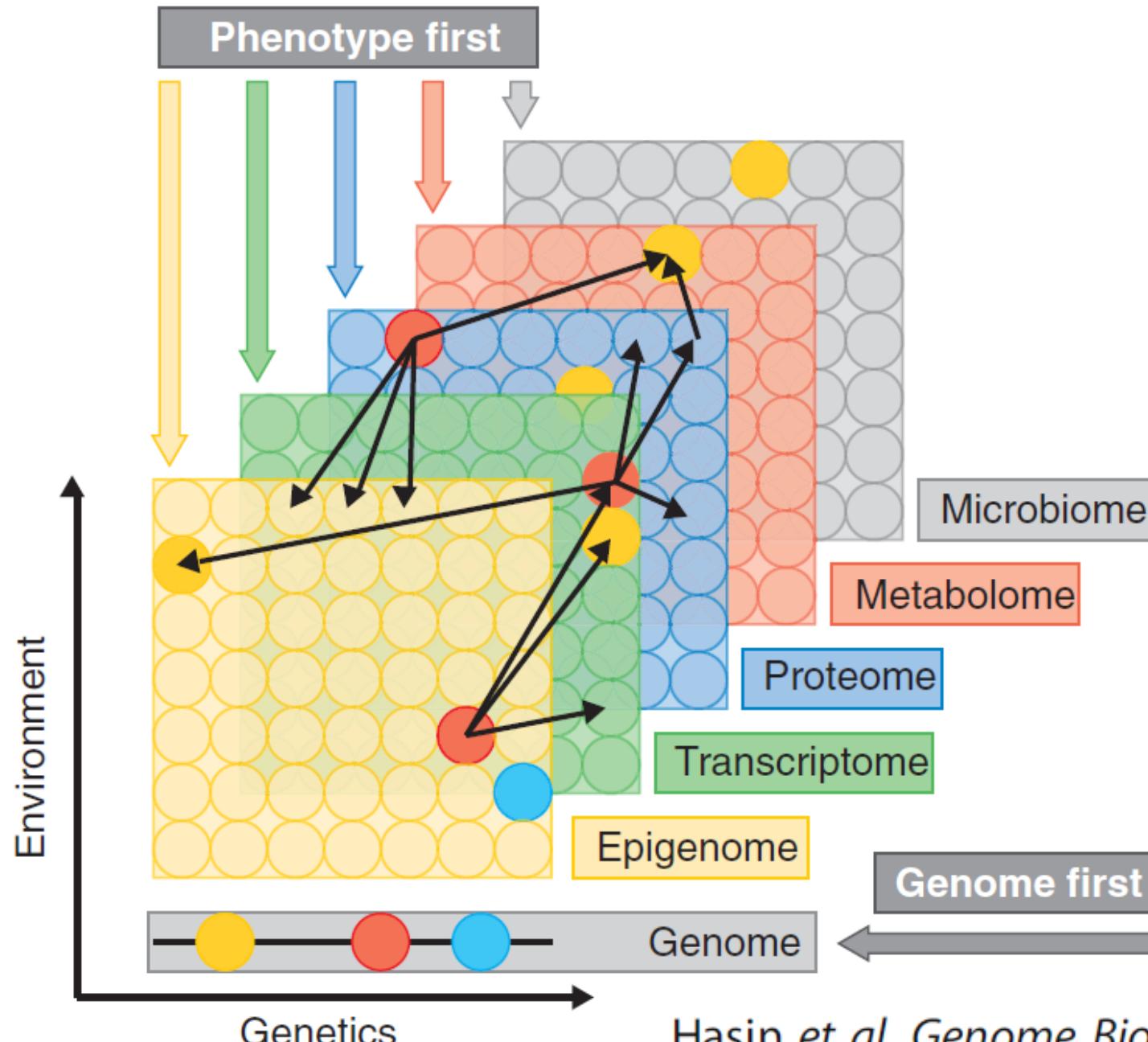
Leading Edge
Review



VIEWPOINT

In the Era of Precision Medicine and Big Data, Who Is Normal?

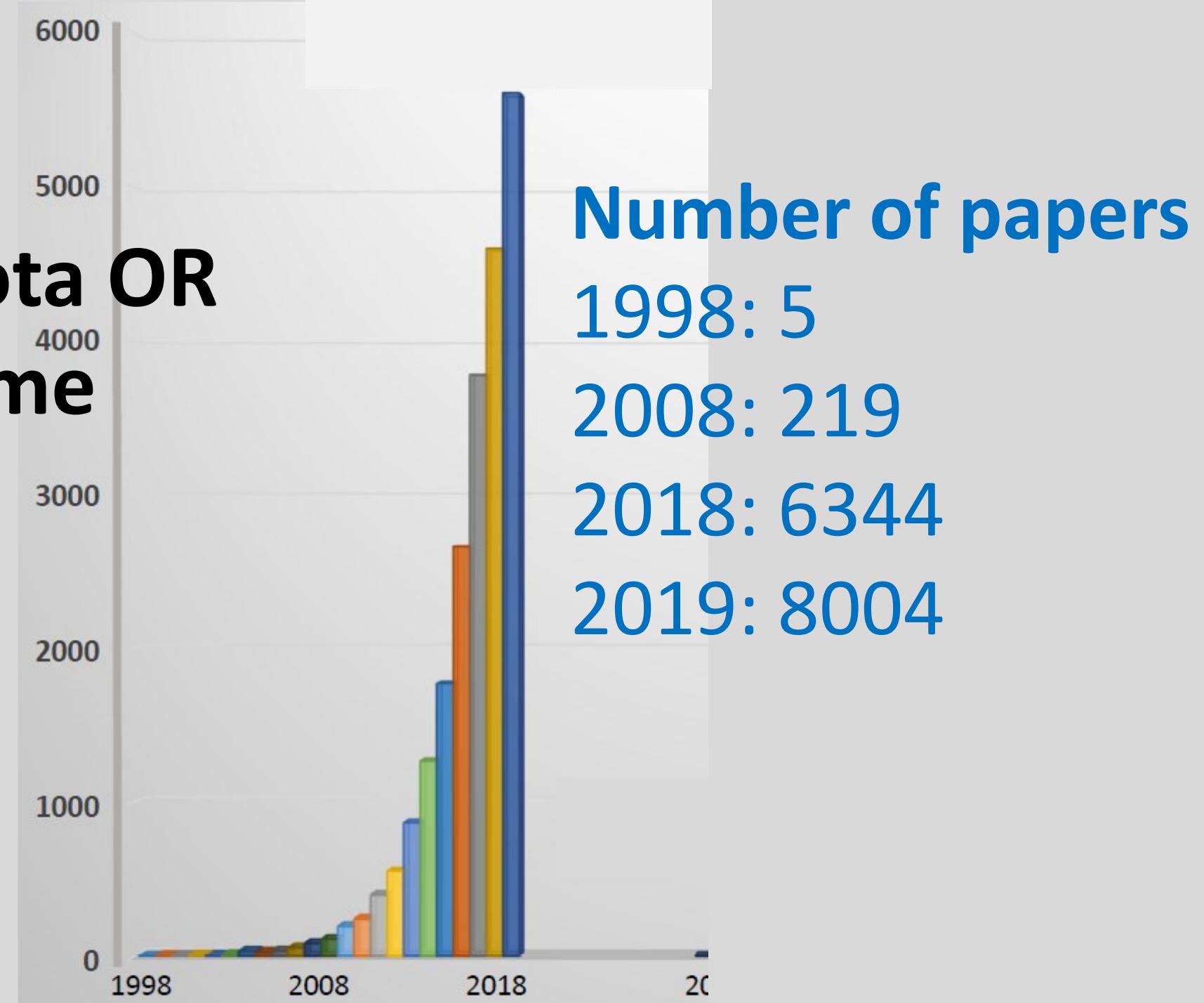
JAMA May 15, 2018 Volume 319, Number 19



**Have you understood
what kind of time we are living
in multi-omics research?**

<https://www.ebi.ac.uk/gwas/docs/diagram-downloads>

PubMed:
"gut microbiota OR
gut microbiome



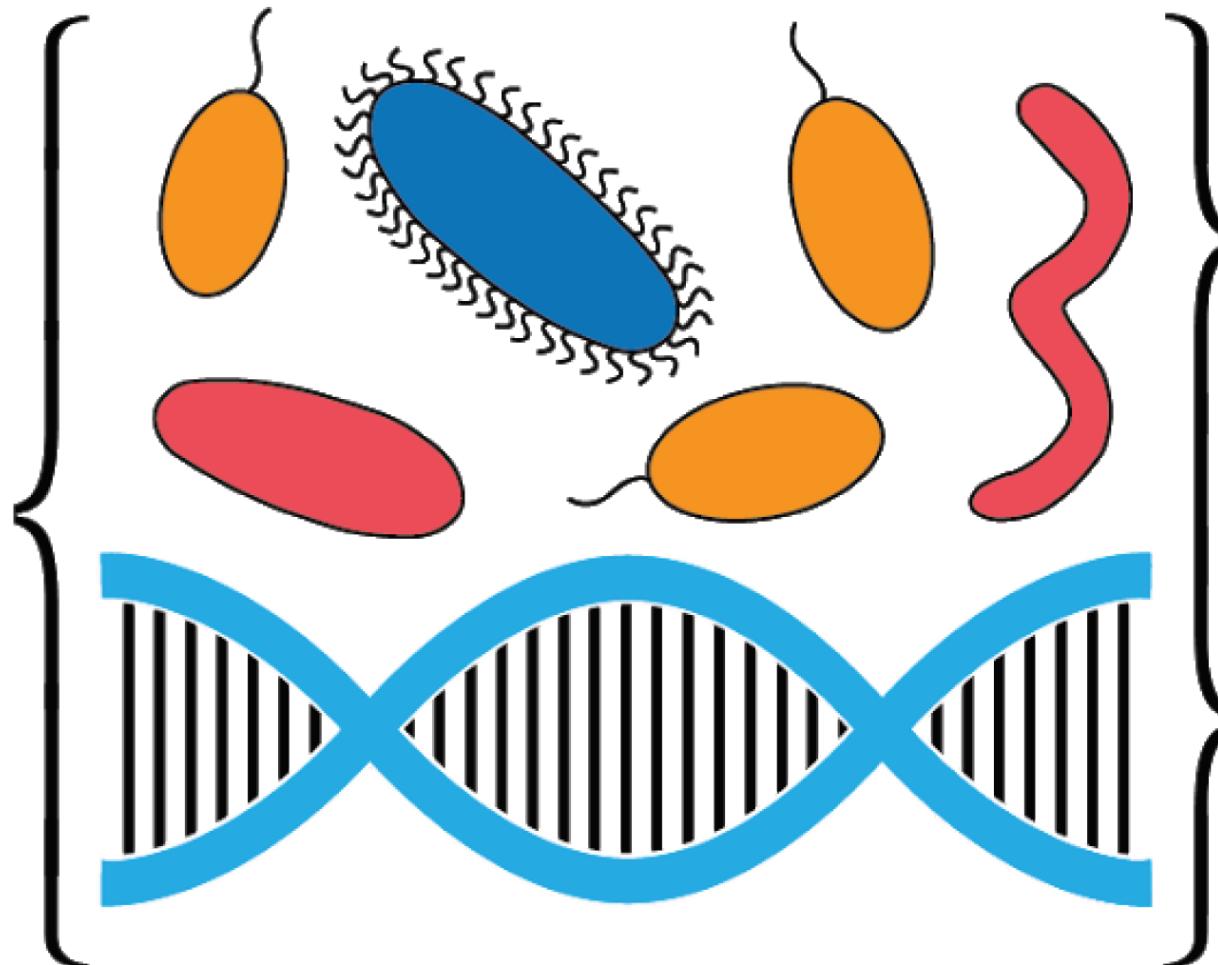
Key differences and similarities between the human genome and metagenome

Complex genetic systems

Play a major role in host health

Can be used to predict host phenotype

Hold potential for novel diagnostics and therapeutics



Semi-transient

>10 million genes

Reflects host environment

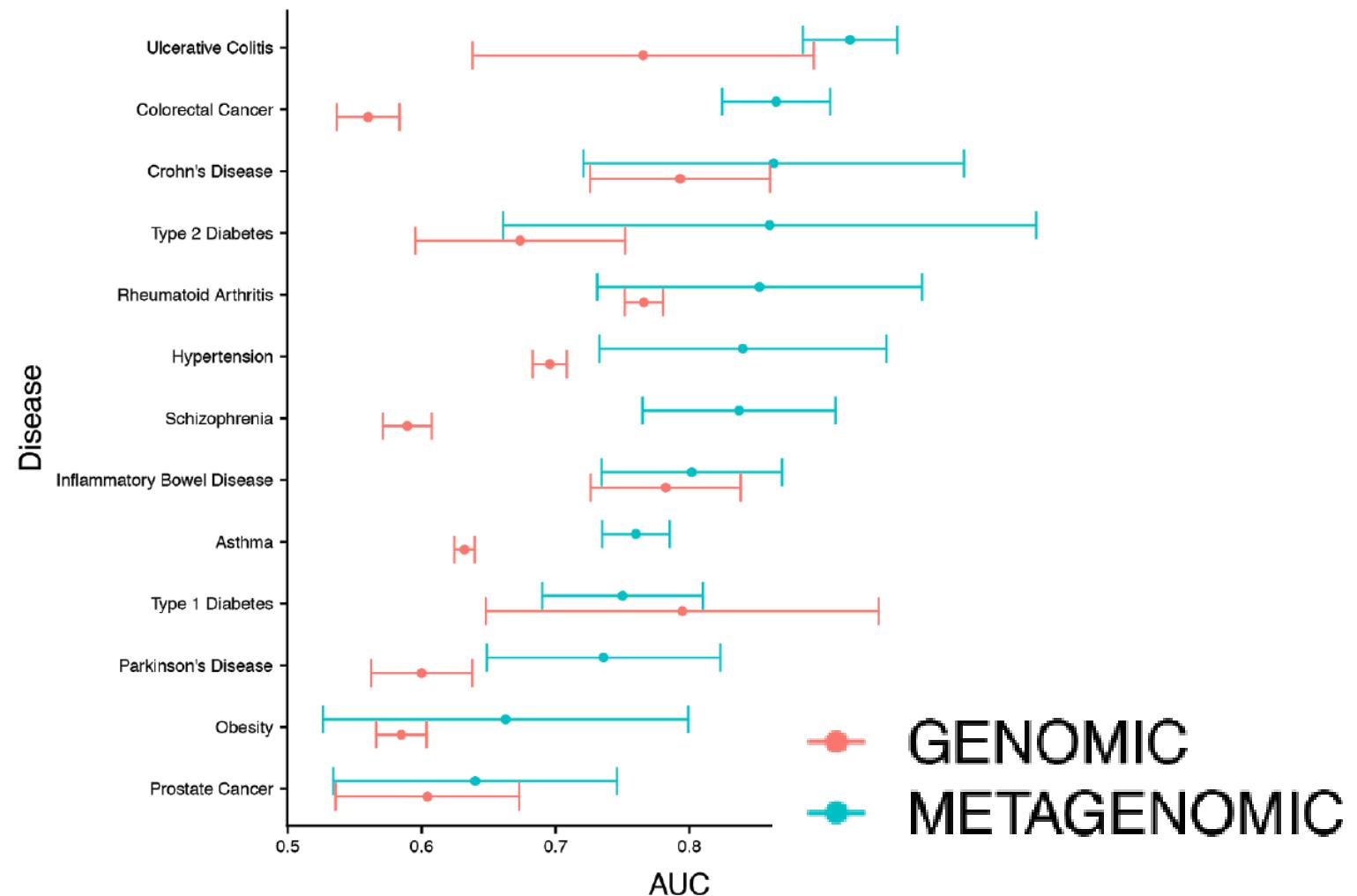
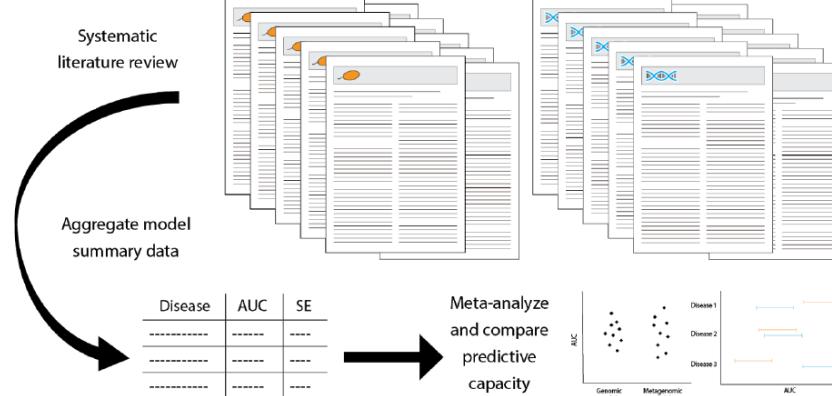
Stable

~20,000 genes

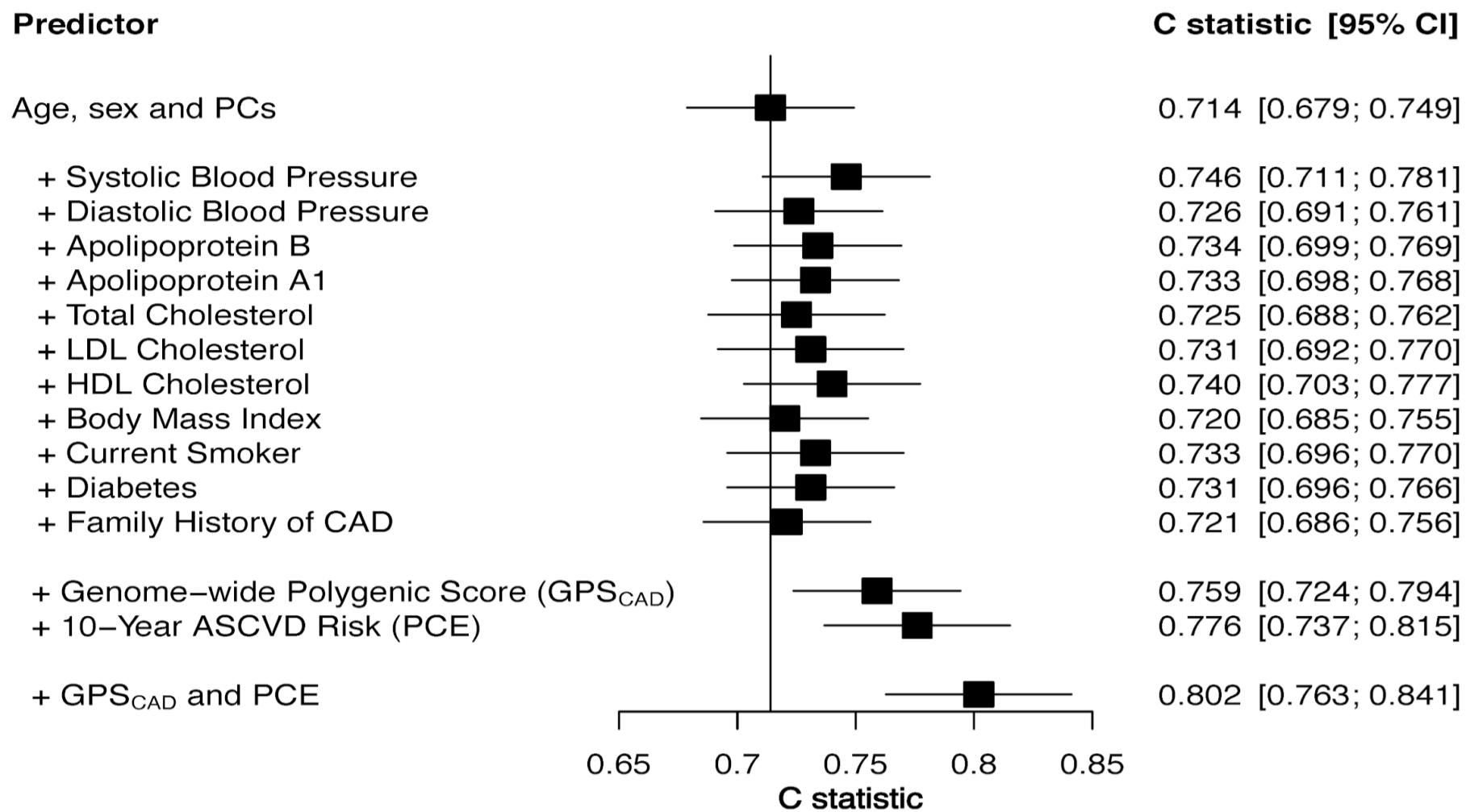
Not environmentally reflective

The predictive power of the microbiome exceeds that of genome-wide association studies in the discrimination of complex human disease

Braden T Tierney^{1,2,3,4}, Yixuan He¹, George M Church^{5,6}, Eran Segal^{7,8}, Aleksandar D Kostic^{2,3,4+}, Chirag J Patel¹⁺



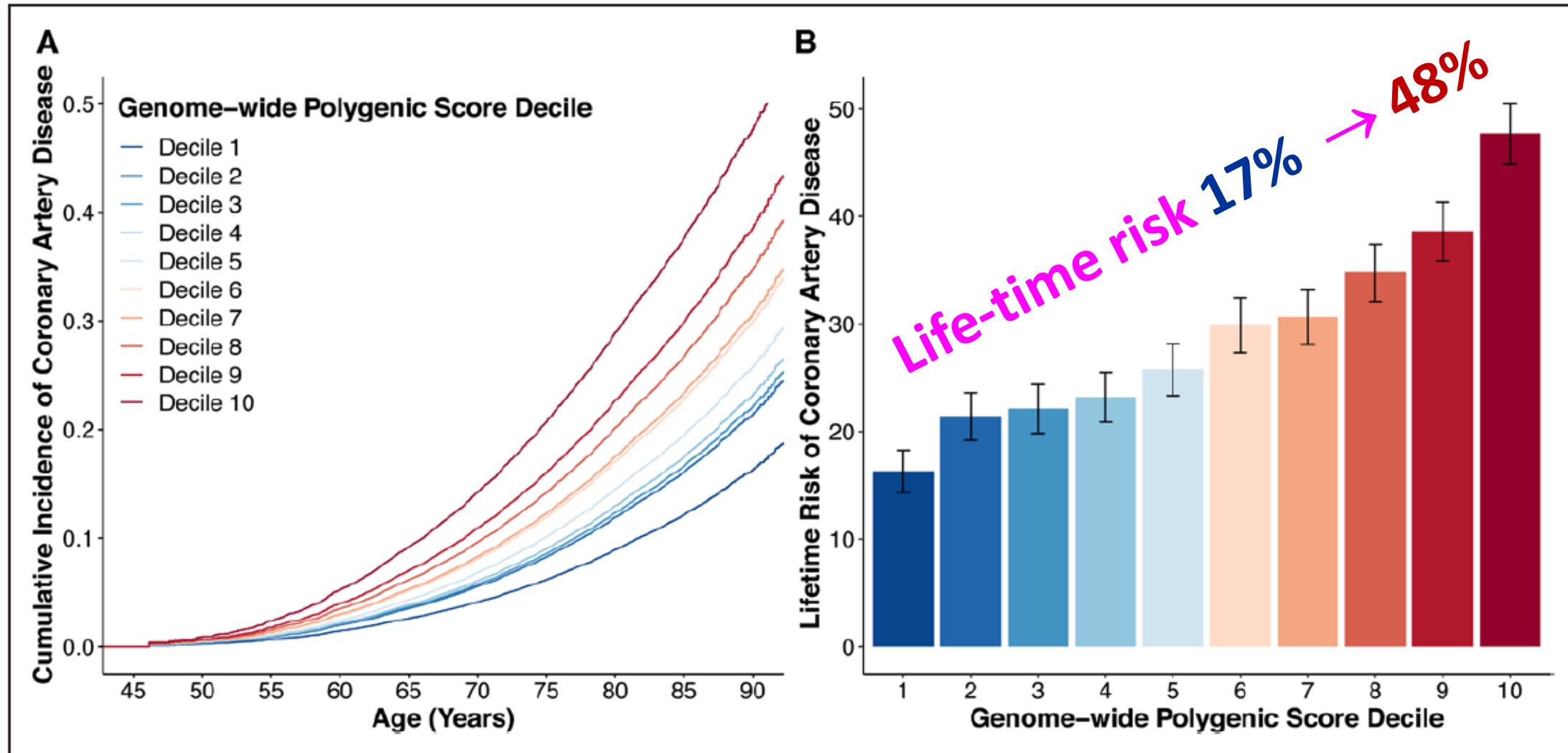
Discriminative capacity of the genome-wide polygenic score and clinical risk factors for coronary artery disease in the Malmö Diet and Cancer Cardiovascular Cohort

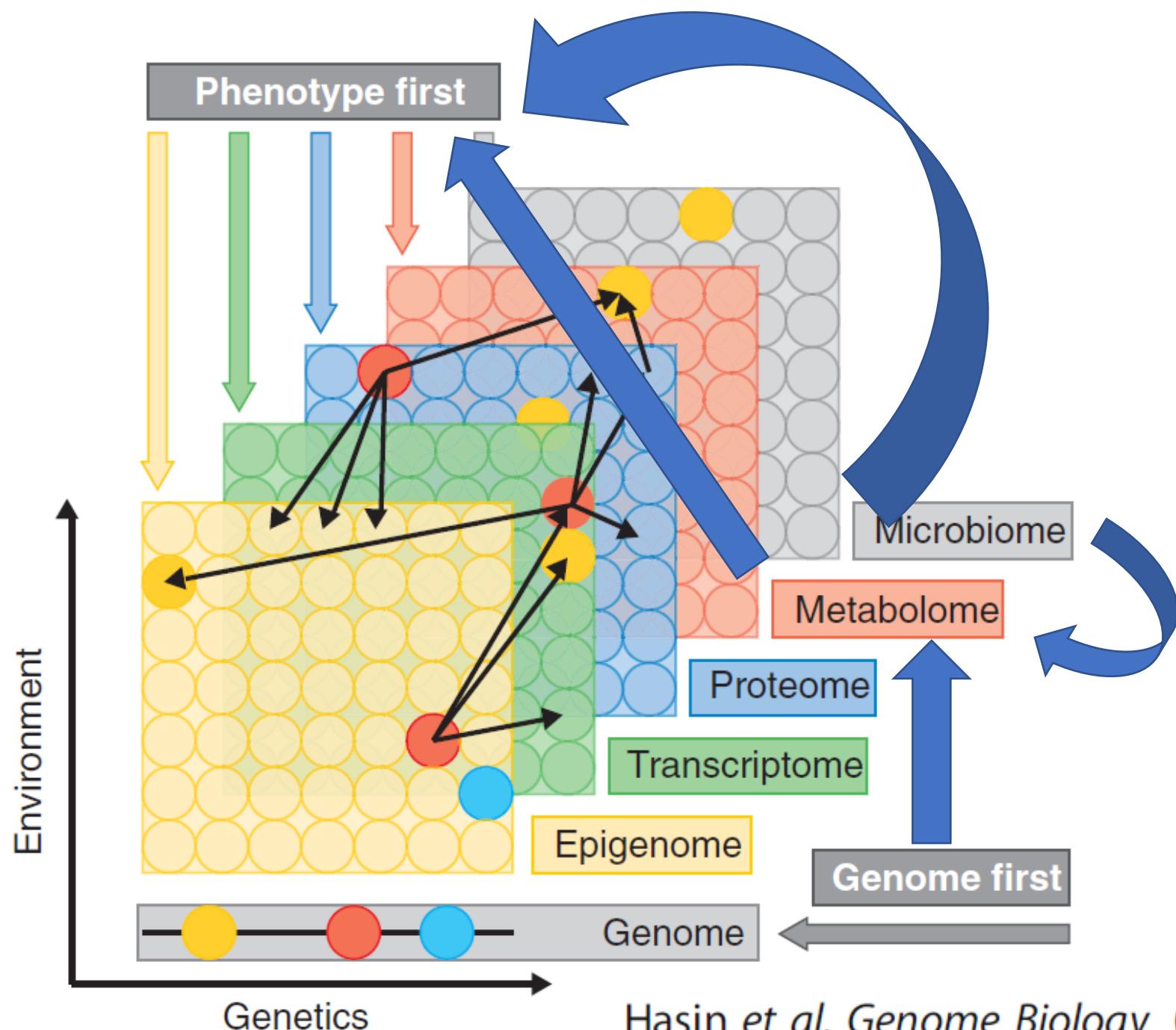


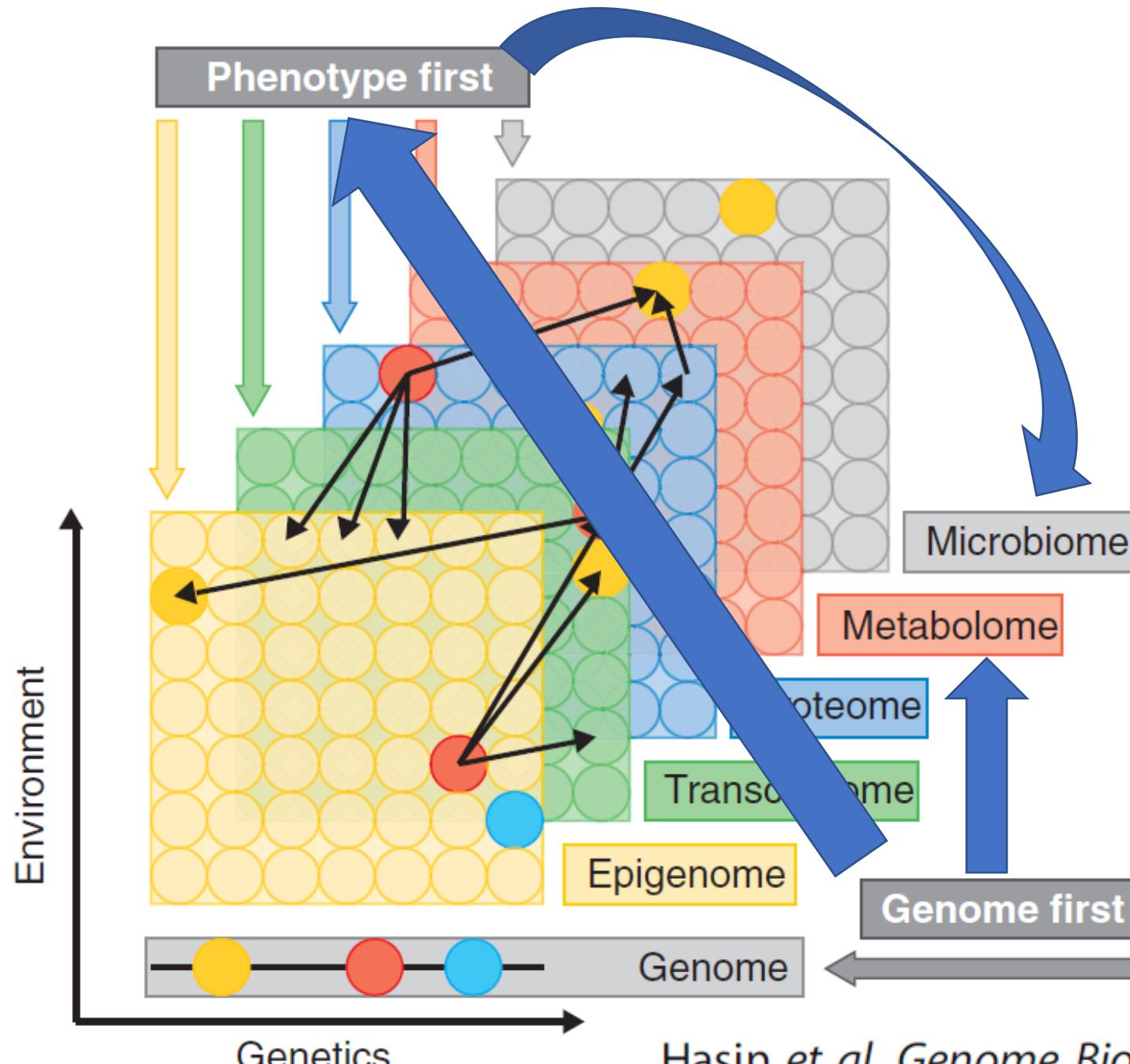
GPS_{CAD} was computed using 6,234,207 SNPs

Hindy et al. ATVB 2020 in press

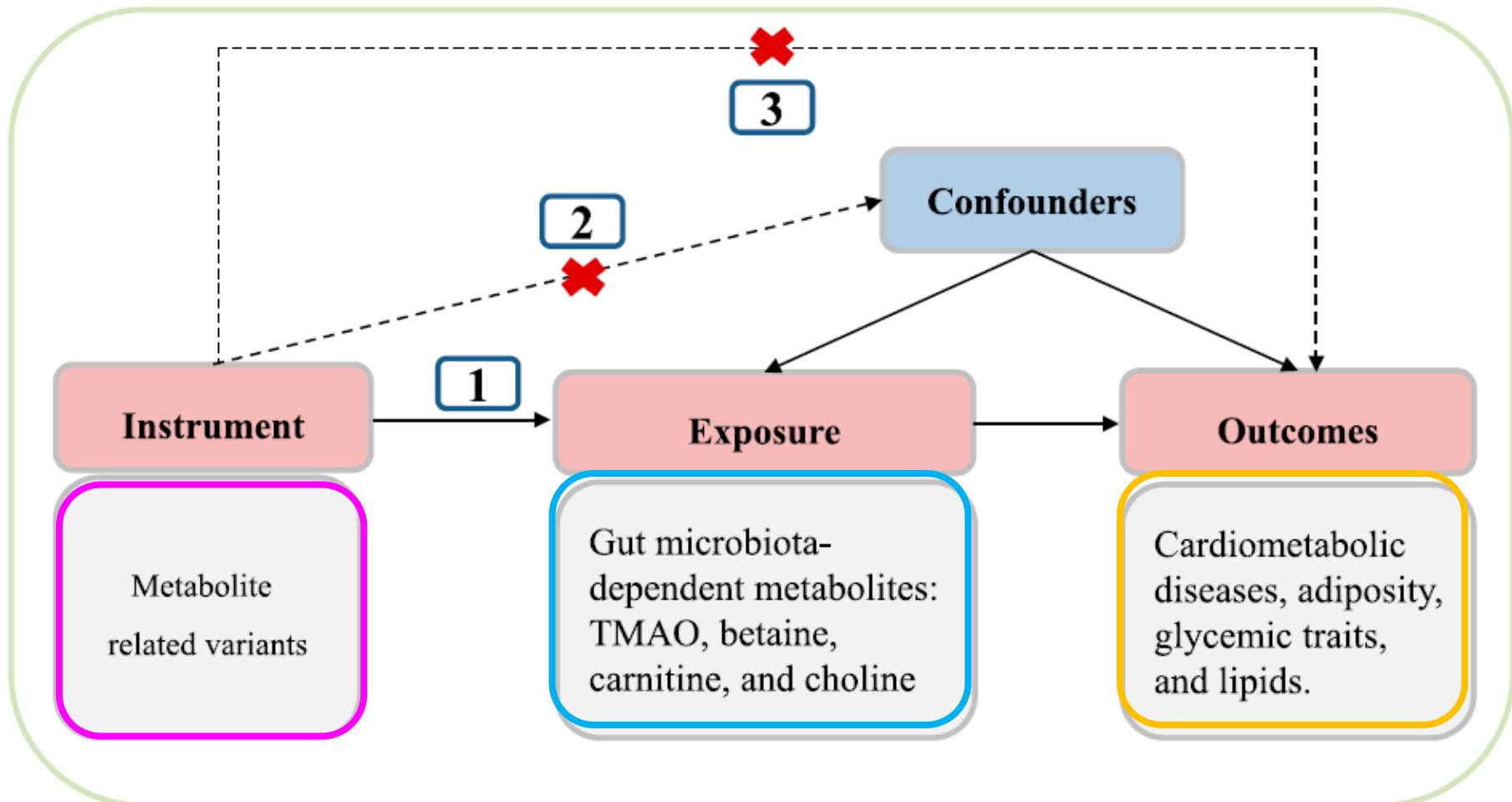
Lifetime risk of coronary artery disease (CAD) in the Malmö Diet and Cancer Study according to decile of the genomewide polygenic score (GPS).



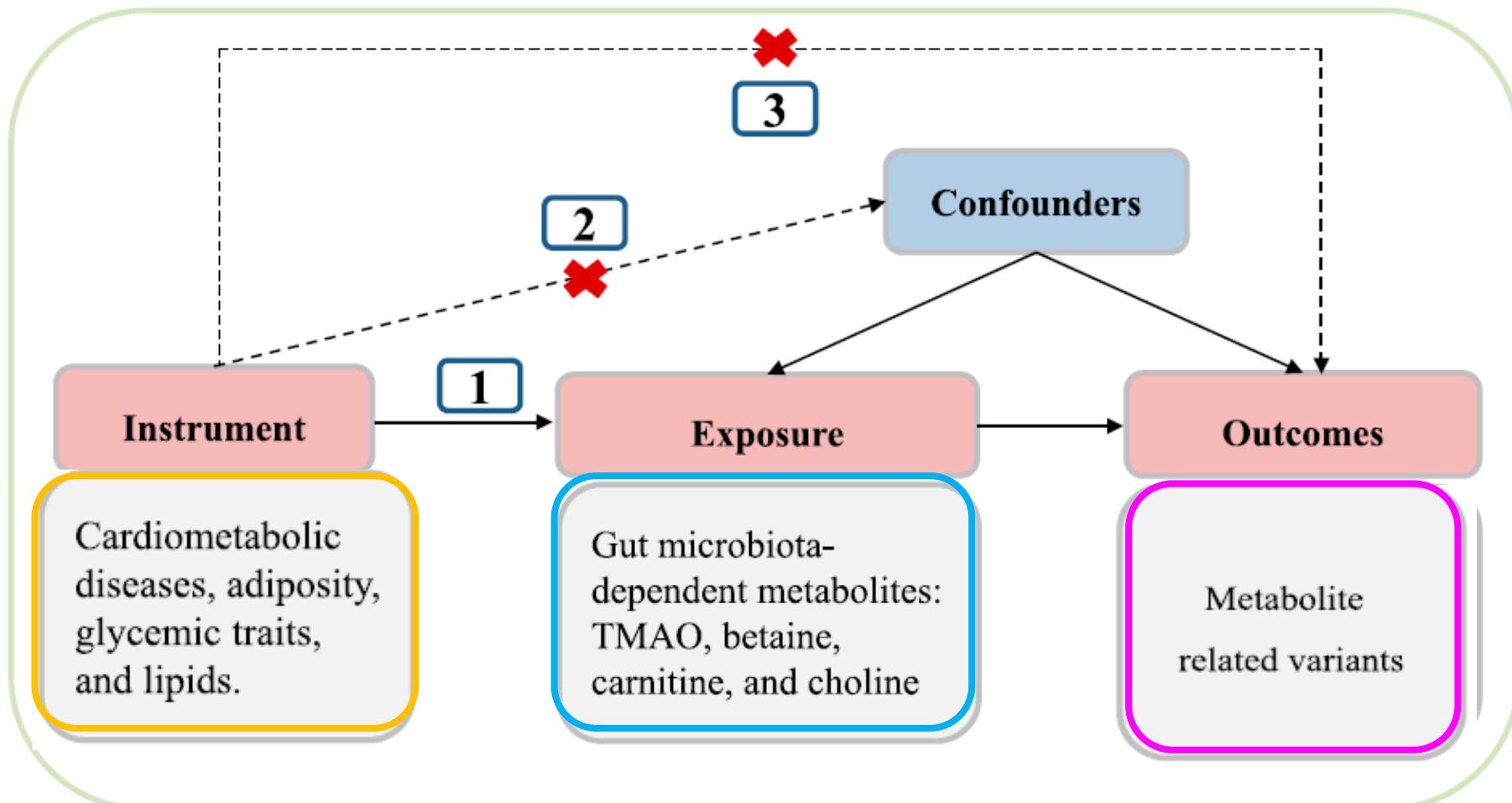




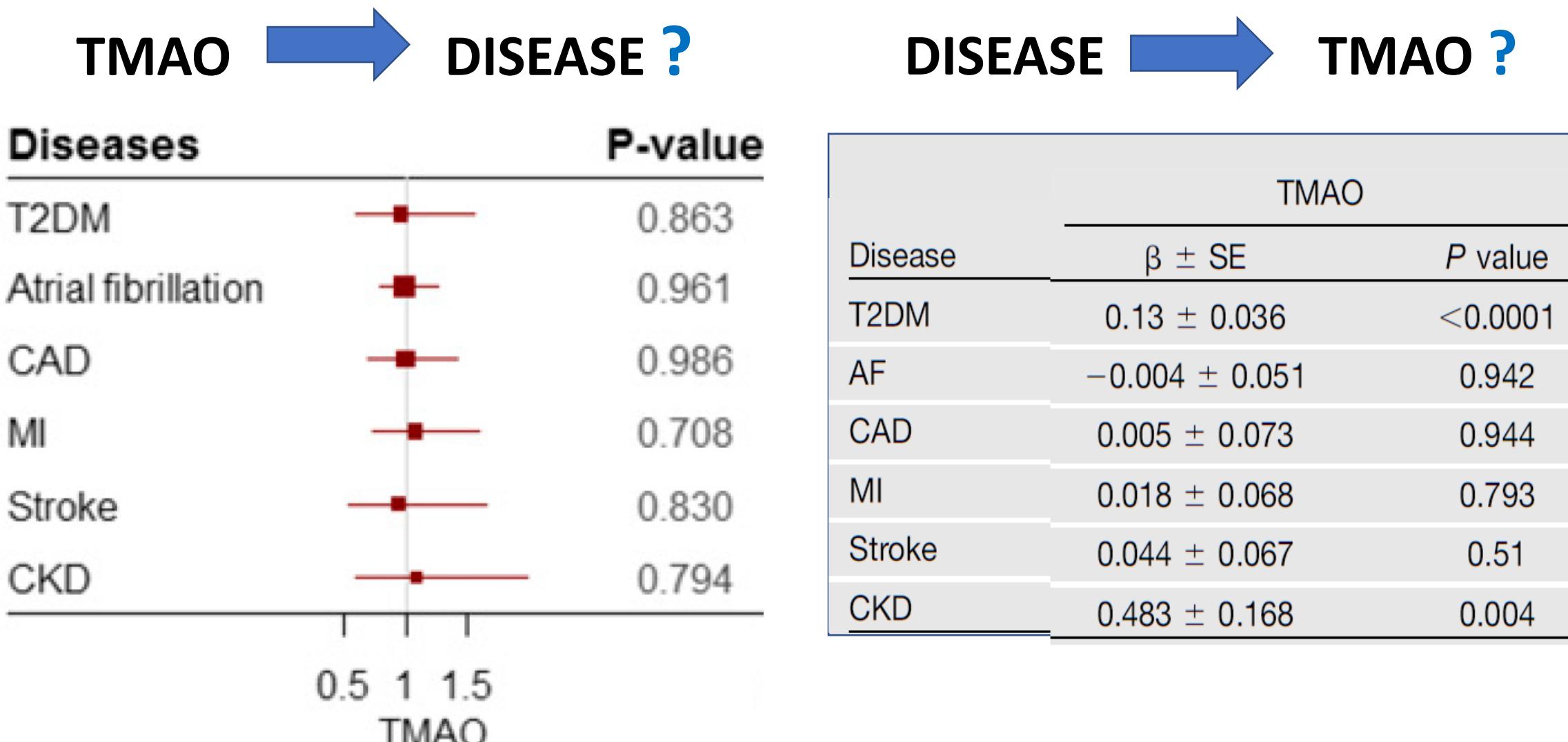
MENDELIAN RANDOMIZATION



MENDELIAN RANDOMIZATION



Causal connection between TMAO and cardiometabolic diseases?



TMAO = Trimethylamine N-oxide

Diabetes 2019;68:1747–1755 | <https://doi.org/10.2337/db19-0153>

<https://doi.org/10.1038/s41467-020-18148-7>

OPEN

Integration of molecular profiles in a longitudinal wellness profiling cohort

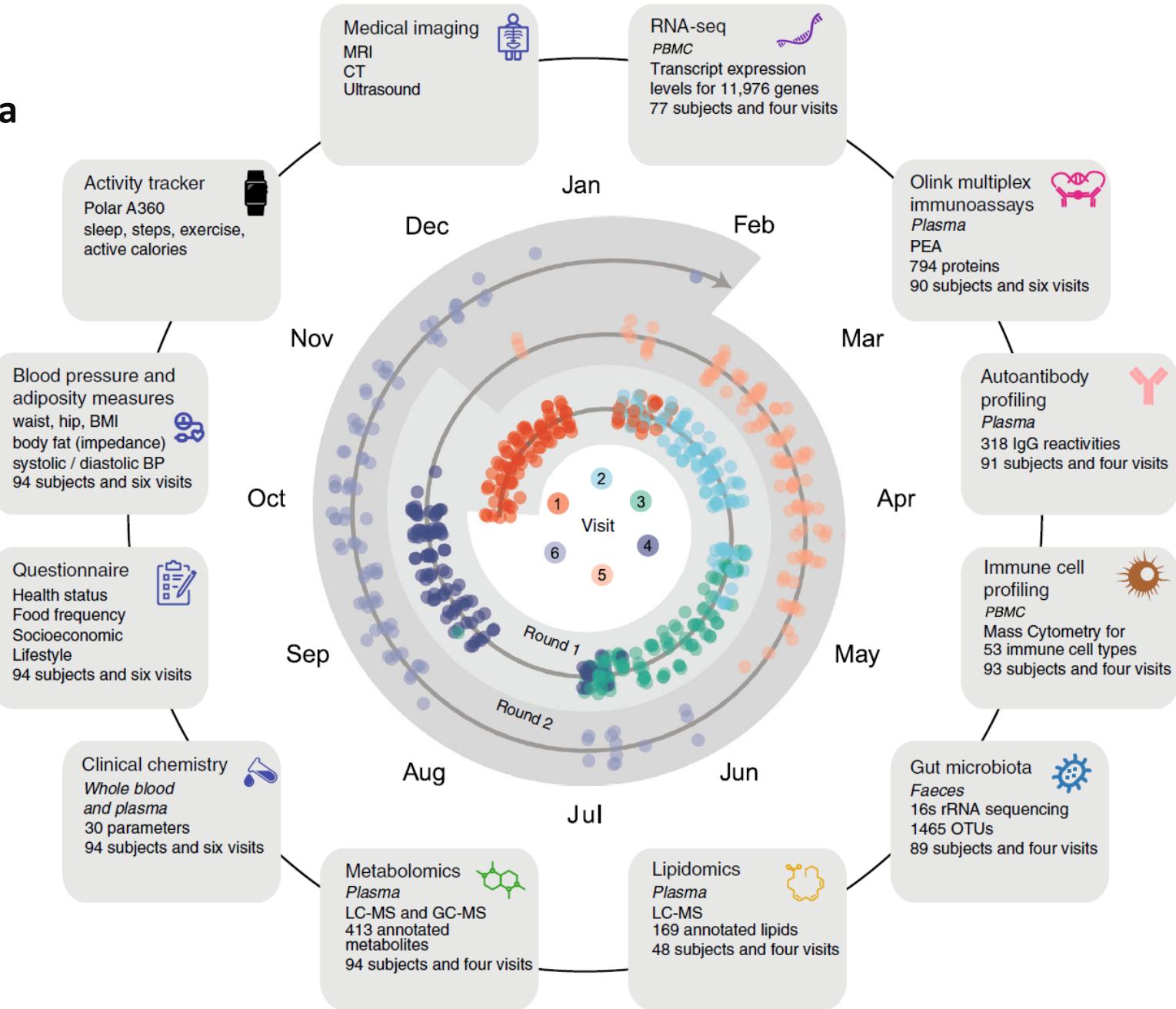
Abdellah Tebani¹, Anders Gummesson^{ID 2,3}, Wen Zhong^{ID 1}, Ina Schuppe Koistinen^{1,4}, Tadepally Lakshmikanth^{ID 5}, Lisa M. Olsson^{ID 2}, Fredrik Boulund^{ID 4}, Maja Neiman¹, Hans Stenlund⁶, Cecilia Hellström^{ID 1}, Max J. Karlsson^{ID 1}, Muhammad Arif^{ID 1}, Tea Dodig-Crnković^{ID 1}, Adil Mardinoglu^{ID 1,7}, Sunjae Lee¹, Cheng Zhang^{ID 1}, Yang Chen⁵, Axel Olin^{ID 5}, Jaromir Mikes^{ID 5}, Hanna Danielsson^{ID 4}, Kalle von Feilitzen¹, Per-Anders Jansson^{2,8}, Oskar Angerås^{9,10}, Mikael Huss^{11,12}, Sanela Kjellqvist¹³, Jacob Odeberg^{ID 1}, Fredrik Edfors^{ID 1}, Valentina Tremaroli², Björn Forsström¹, Jochen M. Schwenk^{ID 1}, Peter Nilsson¹, Thomas Moritz¹⁴, Fredrik Bäckhed^{ID 2,15,16}, Lars Engstrand⁴, Petter Brodin⁵, Göran Bergström^{ID 2,15}, Mathias Uhlen^{ID 1,17} & Linn Fagerberg^{ID 1✉}

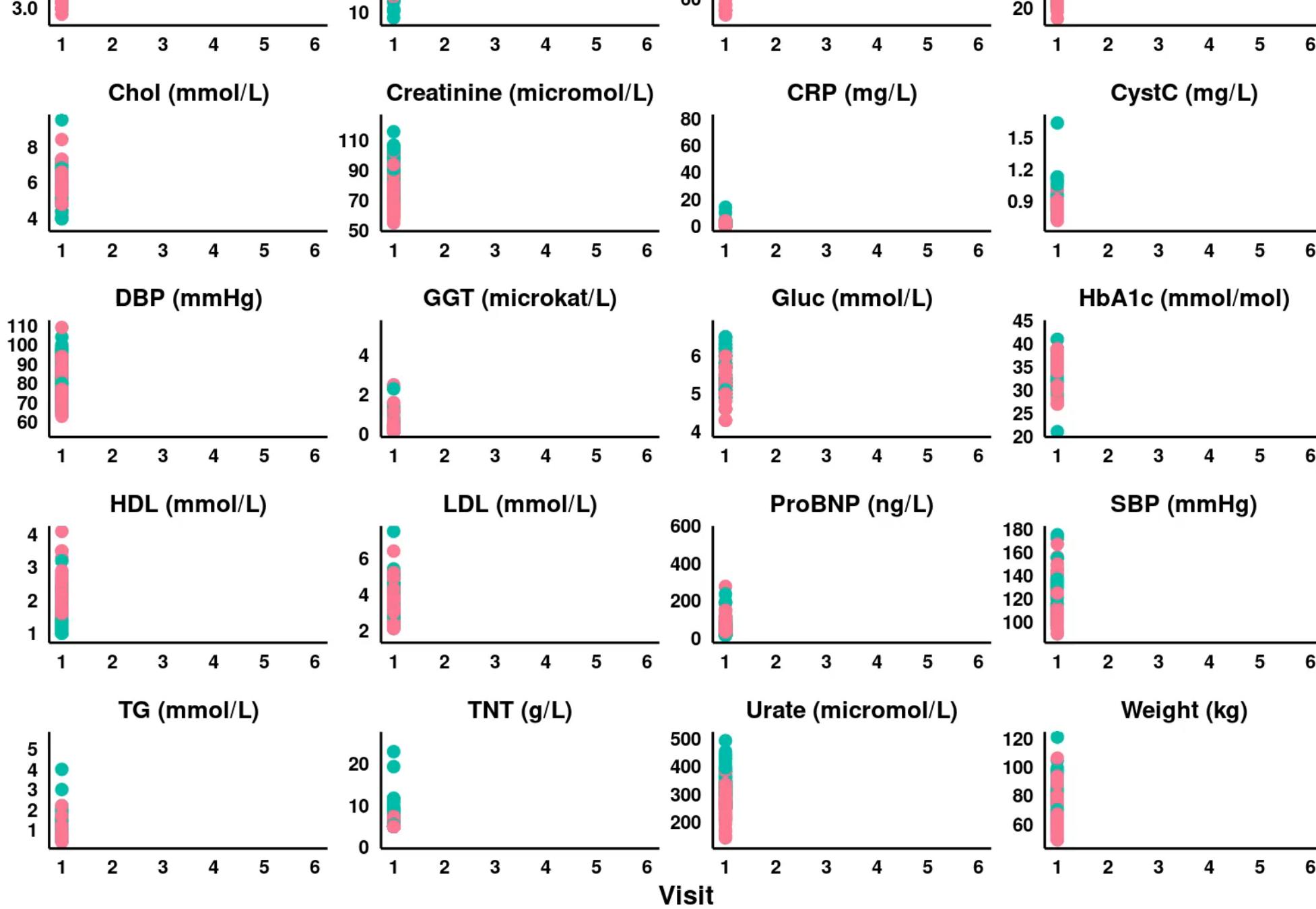
SCAPIS sub-study: 101 individuals (50-65y) between 50 and 65 years profiled by a combination of

- Classical clinical chemistry
- Advanced medical imaging
- Extensive omics:
 - plasma proteome
 - plasma metabolome
 - blood cell composition
 - Transcriptome
 - autoantibody reactivity
 - gut microbiota

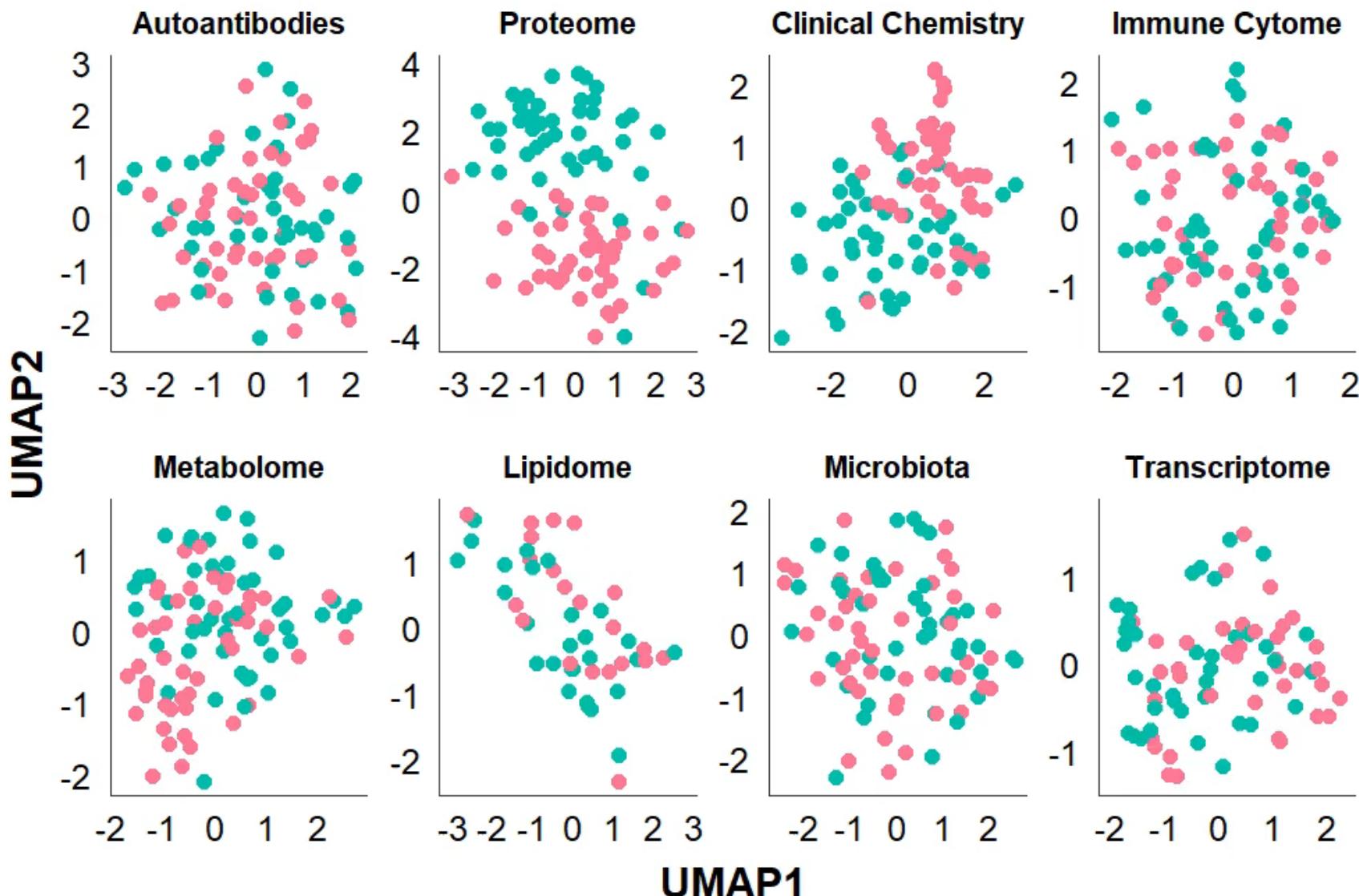
AIMS:

- To probe the uniqueness and stability of an individual's molecular profiles during a 2-year period (6 time points)
- Investigate the relationships between omics profiles and classical routine clinical chemistry measurements.





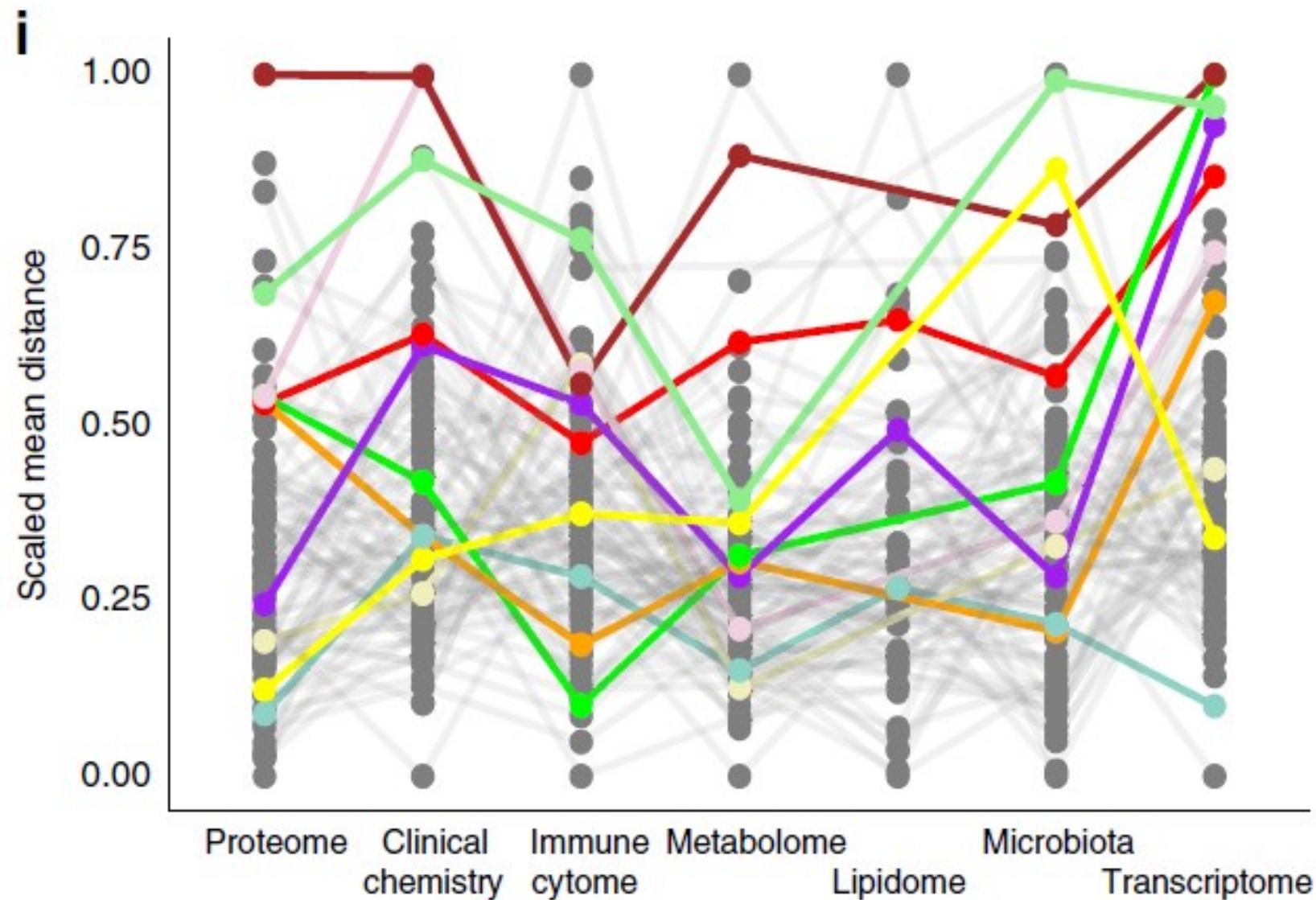
Analyzing the individual longitudinal stability and variability.

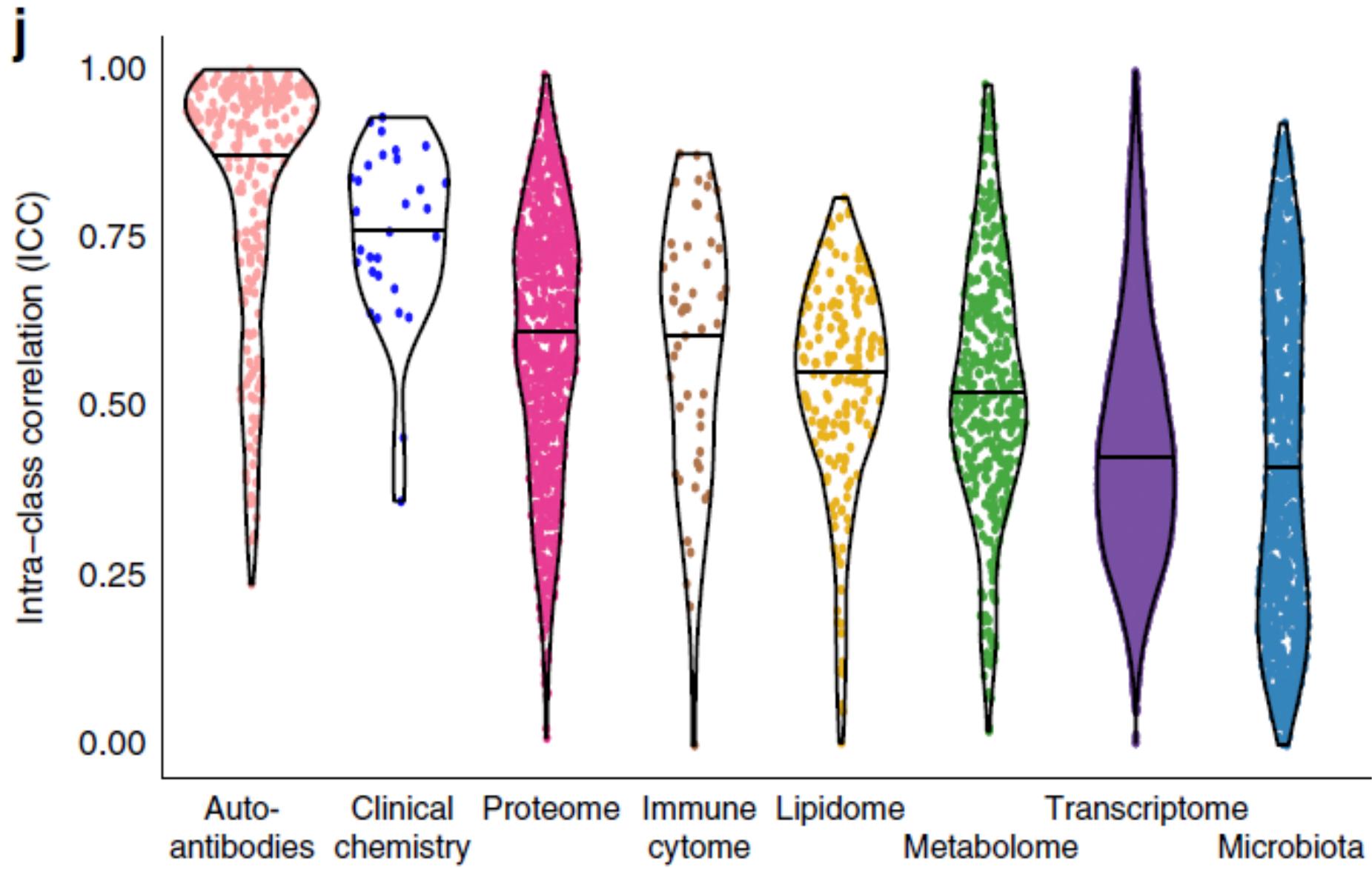


The integrated molecular profiles of all individuals across all analyzed visits are visualized using two-dimensional maps generated by the dimension reduction technique Uniform Manifold Approximation and Projection (UMAP).

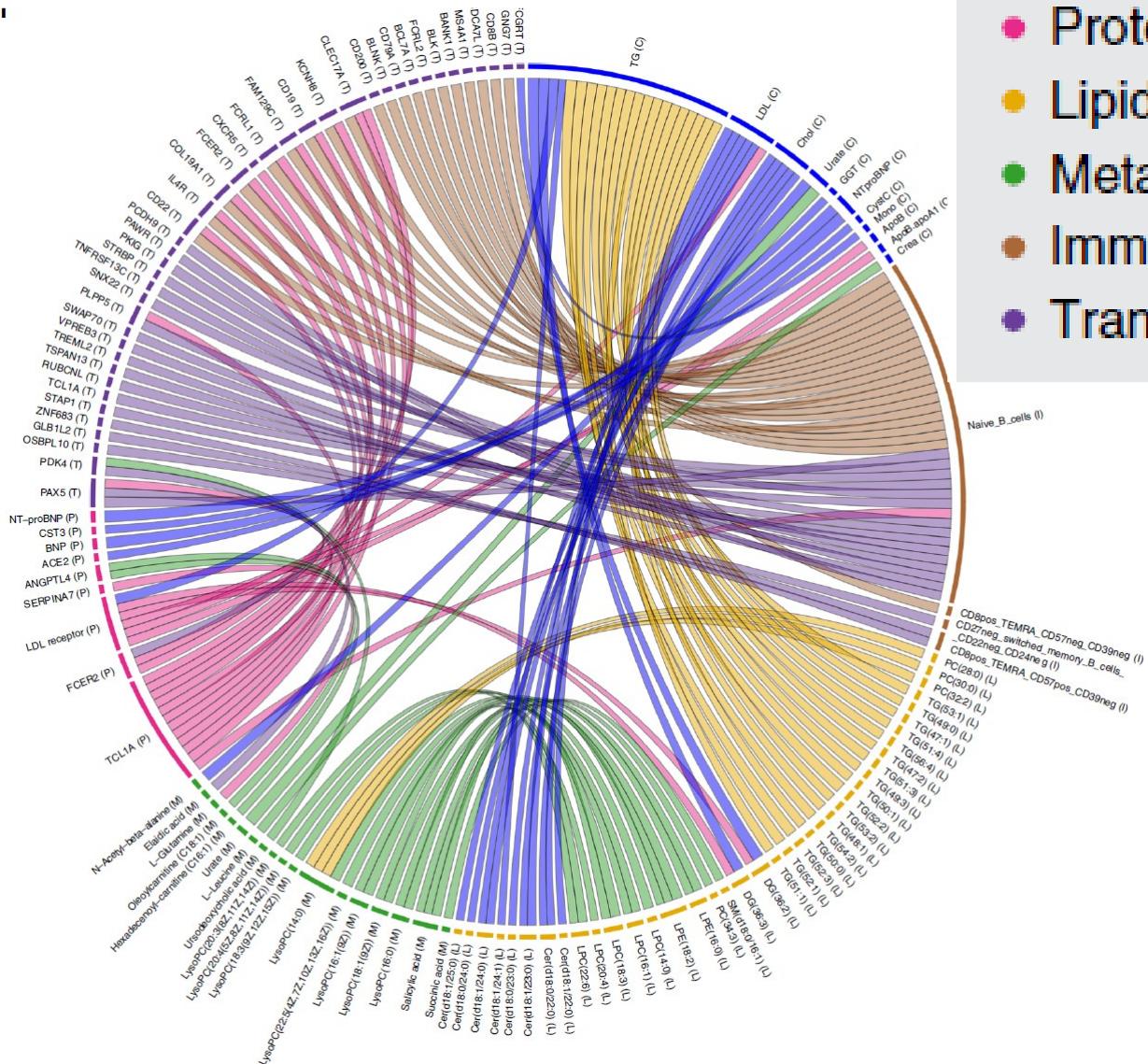
Variability for the individuals across different omics datasets

visualizes the average of the distances for all individuals

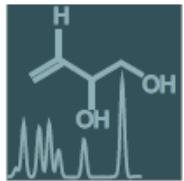




The results support an individual-based definition of health and show that comprehensive omics profiling in a longitudinal manner is a path forward for precision medicine.



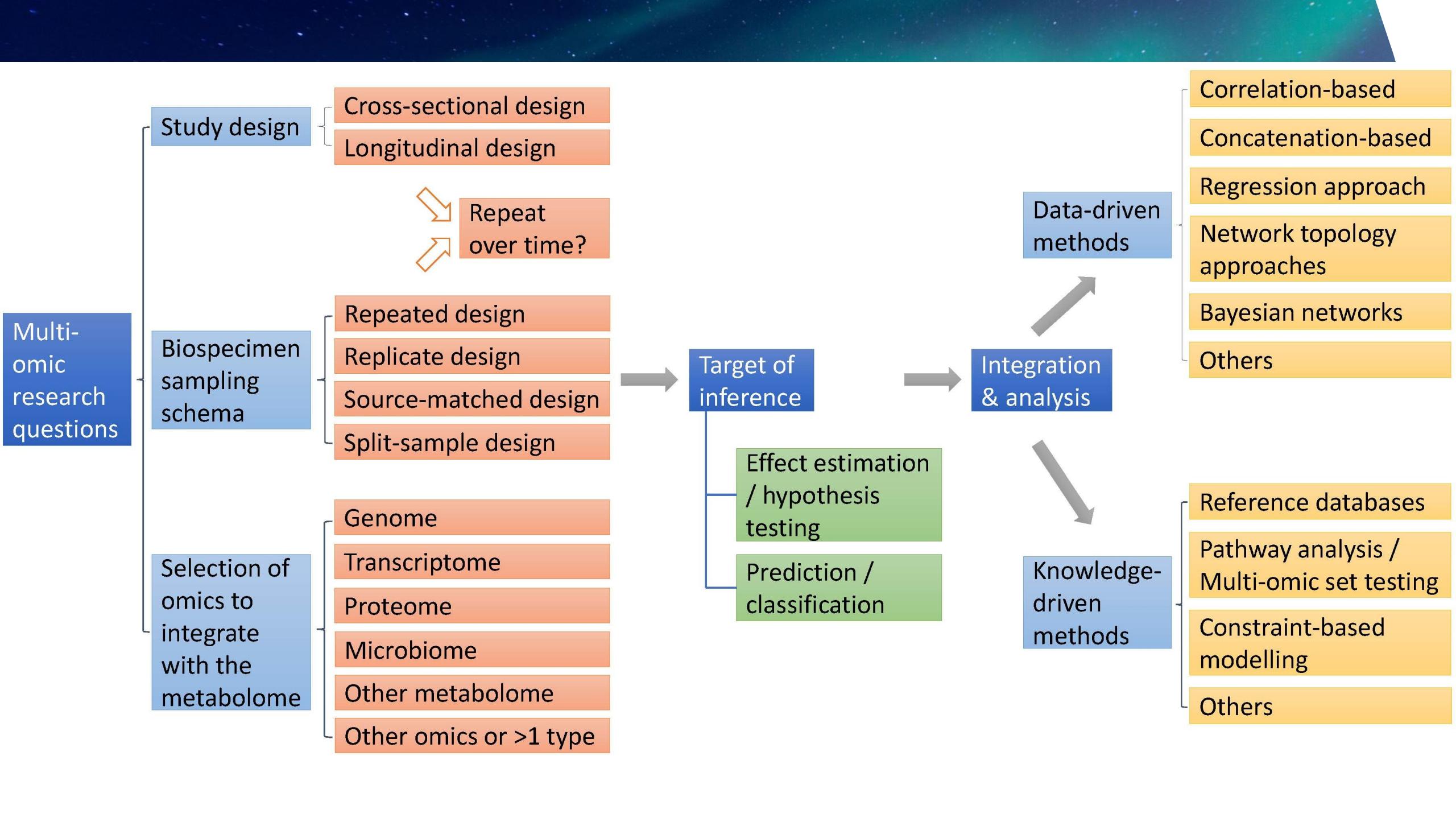
- Clinical chemistry (C)
- Proteome (P)
- Lipidome (L)
- Metabolome (M)
- Immune cytome (I)
- Transcriptome (T)



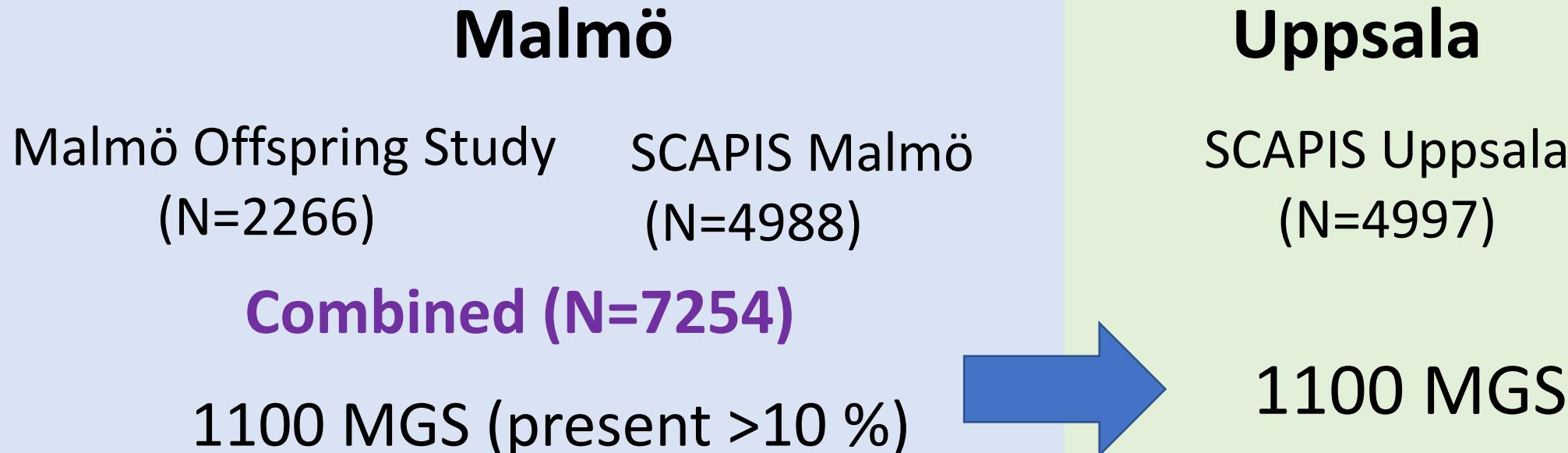
Review

Integration of Metabolomic and Other Omics Data in Population-Based Study Designs: An Epidemiological Perspective

Su H. Chu ^{1,†,*}, Mengna Huang ^{1,†}, Rachel S. Kelly ¹, Elisa Benedetti ², Jalal K. Siddiqui ³, Oana A. Zelezniak ¹, Alexandre Pereira ⁴, David Herrington ⁵, Craig E. Wheelock ⁶, Jan Krumsiek ², Michael McGeachie ¹, Steven C. Moore ⁷, Peter Kraft ⁸, Ewy Mathé ³, Jessica Lasky-Su ^{1,†} and on behalf of the Consortium of Metabolomics Studies Statistics Working Group



Gut microbiota in >12,000 Swedes



- Deep cardiometabolic phenotyping
- Register based follow-up
- Follow-up visits?

All cohorts N=12,251
+ Metabolomic data N= 11,263
+ Genetic data
+ Proteomic data



clinical
microbiomics

H. Bjørn Nielsen, PhD
Clinical Microbiomics
Jacob Bak Holm, PhD
Clinical Microbiomics

Based on binning of co-abundant genes across the metagenomic samples

Discovery and quantification of all bacteria and virus with very high precision and coverage

Aids microbial genome assembly without need for reference sequences

Relies on the notion that abundance is constant across genetic entities (=each gene on a specific bacteria chromosome will be found in a sample in the same abundance as any other gene on that chromosome)

ARTICLES

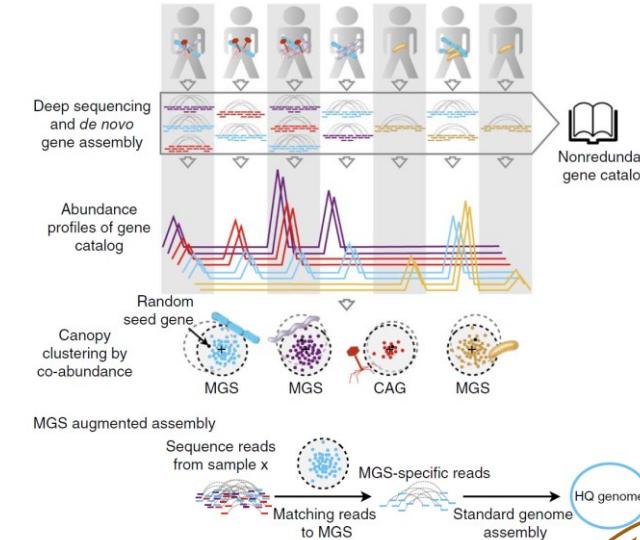
nature
biotechnology

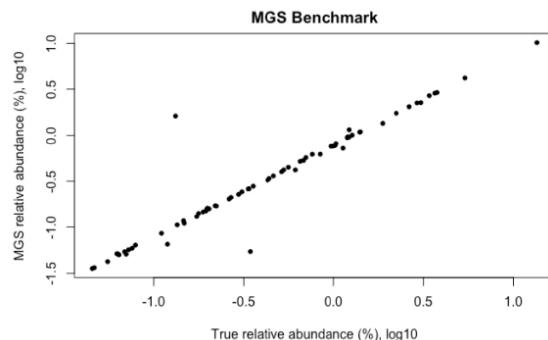
2014

Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes

H Bjørn Nielsen^{1,2,32}, Mathieu Almeida^{3–5,32}, Agnieszka Sierakowska Juncker^{1,2}, Simon Rasmussen¹, Junhua Li^{6–8}, Shinichi Sunagawa⁹, Damian R Plichta¹, Laurent Gautier¹, Anders G Pedersen¹, Emmanuelle Le Chatelier^{3,4}, Eric Pelletier^{10–12}, Ida Bonde^{1,2}, Trine Nielsen¹³, Chaysavanh Manichanh¹⁴, Manimozhiyan Arumugam^{7,9,13}, Jean-Michel Batto^{3,4}, Marcelo B Quintanilha dos Santos¹, Nikolaj Blom², Natalia Borruel¹⁴, Kristoffer S Burgdorf¹³, Fouad Boumezbeur^{3,4}, Frances Casellas¹⁴, Joël Doré^{3,4}, Piotr Dworzynski¹, Francisco Guarner¹⁴, Torben Hansen^{13,15}, Falk Hildebrand^{16,17}, Rolf S Kaae¹⁸, Sean Kennedy^{3,4}, Karsten Kristiansen^{7,19}, Jens Roat Kultima⁹, Pierre Léonard^{3,4}, Florence Levenez^{3,4}, Ole Lund¹, Bouziane Moumen^{3,4}, Denis Le Paslier^{10–12}, Nicolas Pons^{3,4}, Oluf Pedersen^{13,20–22}, Edi Prifti^{3,4}, Junjie Qin^{6,7}, Jeroen Raes^{17,23,24}, Soren Sorensen²⁵, Julien Tap⁹, Sebastian Tims²⁶, David W Ussery¹, Takiji Yamada^{9,27}, MetaHIT Consortium²⁸, Pierre Renault³, Thomas Sicheritz-Ponten^{1,2}, Peer Bork^{9,29}, Jun Wang^{7,13,19,30}, Soren Brunak^{1,2} & S Dusko Ehrlich^{3,4,31}

Most current approaches for analyzing metagenomic data rely on comparisons to reference genomes, but the microbial diversity of many environments extends far beyond what is covered by reference databases. *De novo* segregation of complex metagenomic data into specific biological entities, such as particular bacterial strains or viruses, remains a largely unsolved problem. Here we present a method, based on binning co-abundant genes across a series of metagenomic samples, that enables comprehensive discovery of new microbial organisms, viruses and co-inherited genetic entities and aids assembly of microbial genomes without the need for reference sequences. We demonstrate the method on data from 396 human gut microbiome samples and identify 7,381 co-abundance gene groups (CAGs), including 741 metagenomic species (MGS). We use these to assemble 238 high-quality microbial genomes and identify affiliations between MGS and hundreds of viruses or genetic entities. Our method provides the means for comprehensive profiling of the diversity within complex metagenomic samples.





High correlation between relative and absolute abundance

(Figure 8) Relative MGS abundance using Clinical Microbiomics' pipeline as compared to the true relative abundance in the in silico generated mock.

The MGS level resolution is achieved by tracking each MGS strain

Allows SNV-based strain-level resolution (=sub-species resolution)

to a "strain" (SNV) within

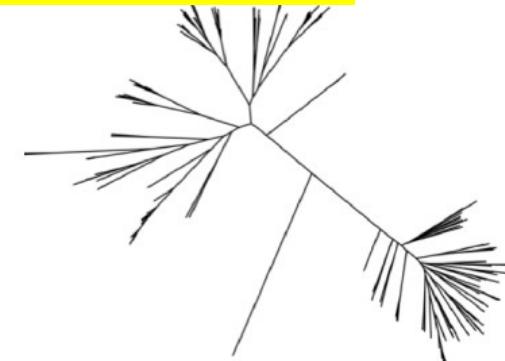
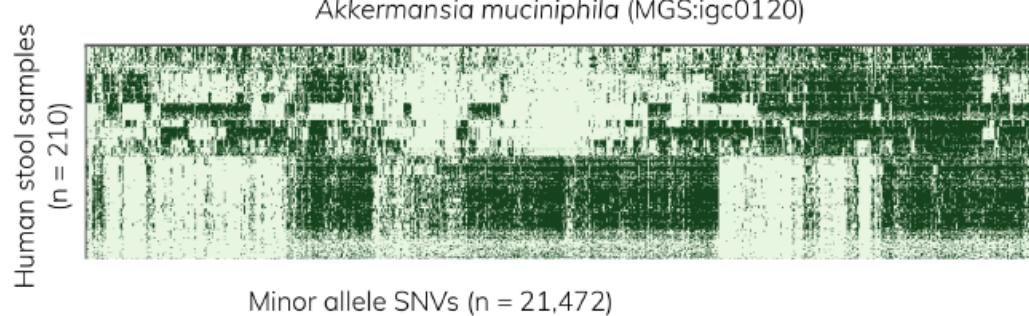
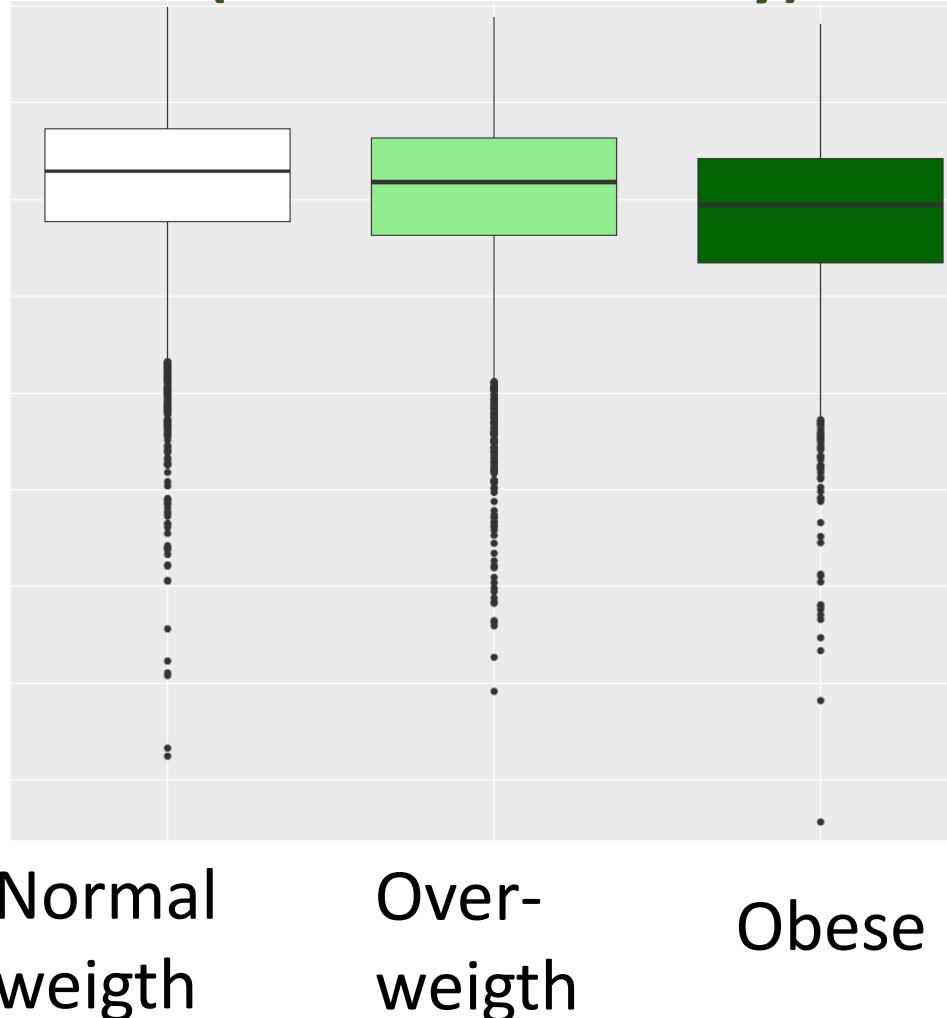


Figure 9) Example of the SNV based sub-species resolution. The data illustrates the SNV profile identified for the *Akkermansia muciniphila* MGS in 210 human stool samples. The SNV-based sub-species populations is visualized with a phylogenetic tree.

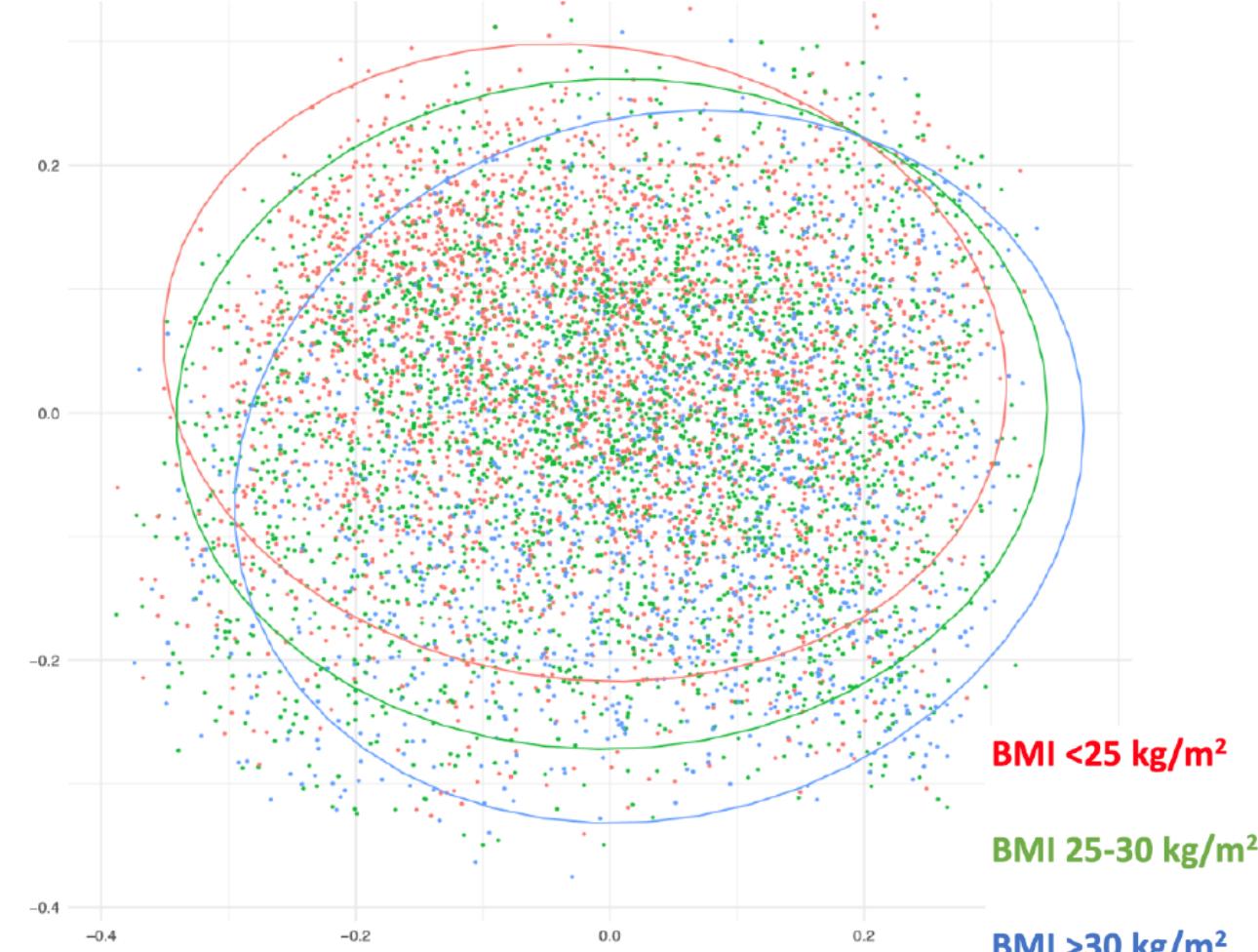


Diversity of gut microbiota by BMI-categories in Malmö cohorts

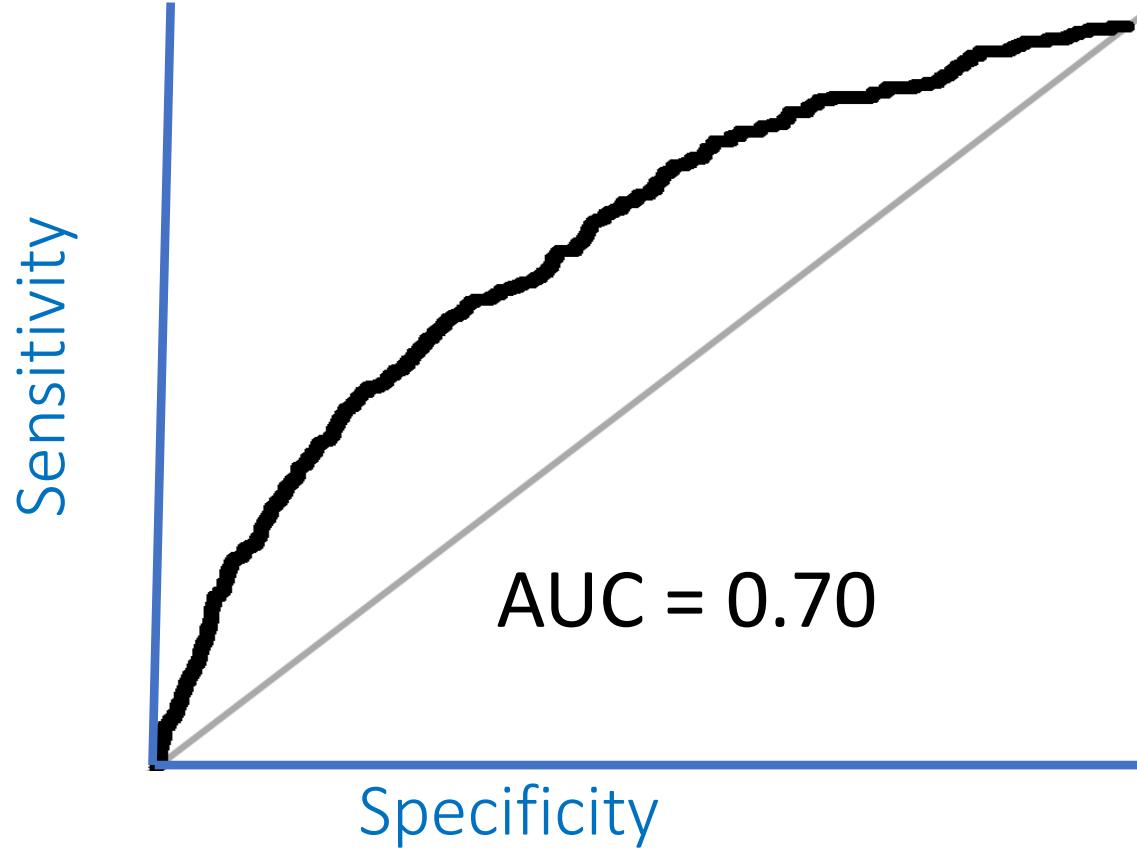
**Alpha-diversity
(Shannon diversity)**



**Beta-diversity
(Bay-Curtis dissimilarity index)**



Beta- diversity and diabetes prediction in MOS ans SCAPIS Malmö (N=7254)



Logistic regression adjusted for alpha-diversity

Metagenomic Species (MGS) and diabetes prediction

MOS +SCAPIS-Mö
N= 7254

TRAINING 80%

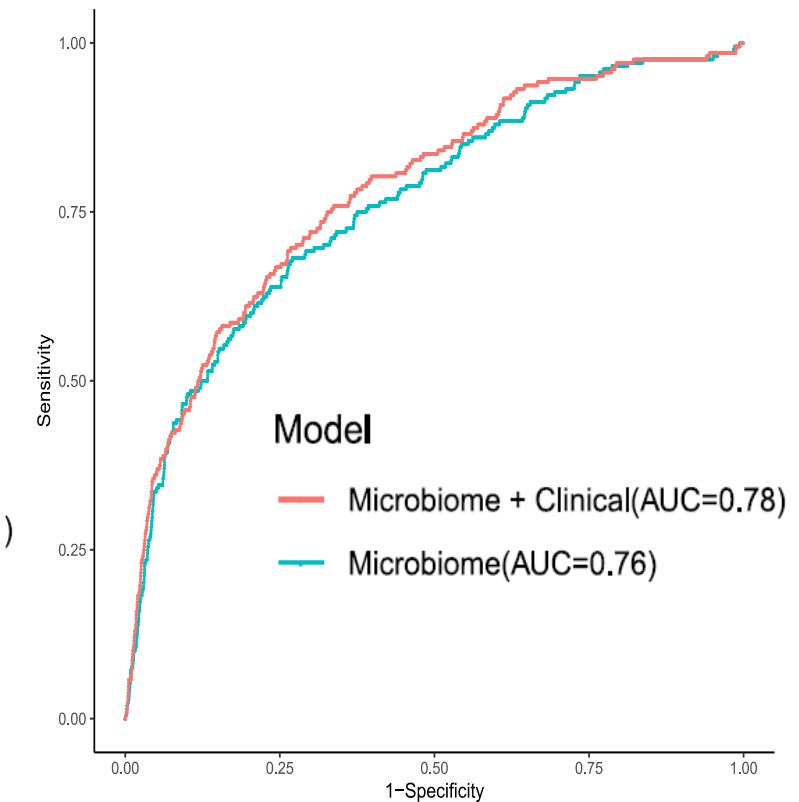
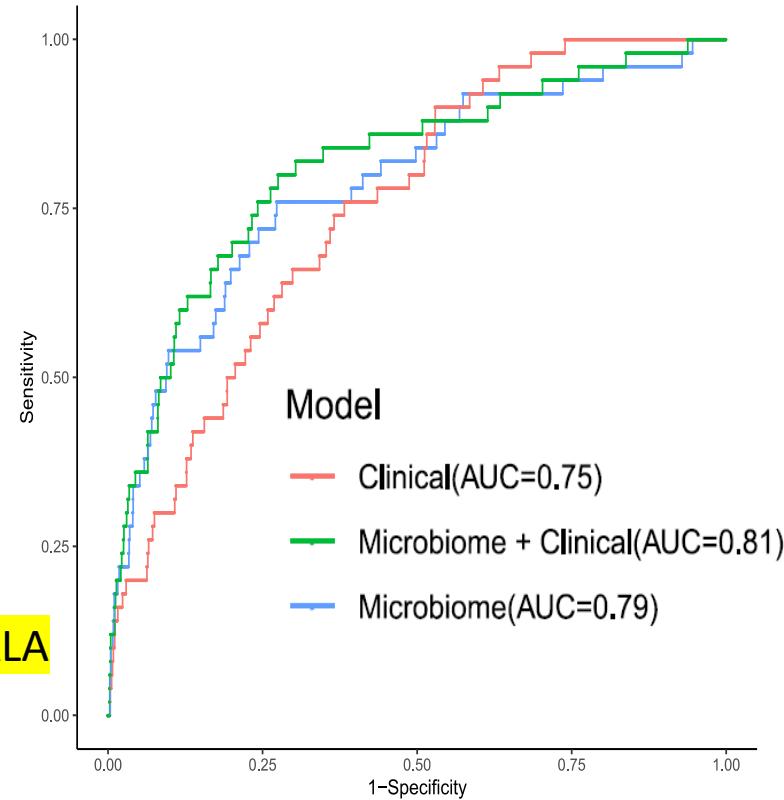
MOS +SCAPIS-Mö
N= 5803

VALIDATION IN
20% MÅLÖ + 100% SCAPIS UPPSALA

N= 1415
SCAPIS-
Uppsala
N=4997

Malmö Cohorts
(VALIDATE 1)

SCAPIS Uppsala
(VALIDATE 2)



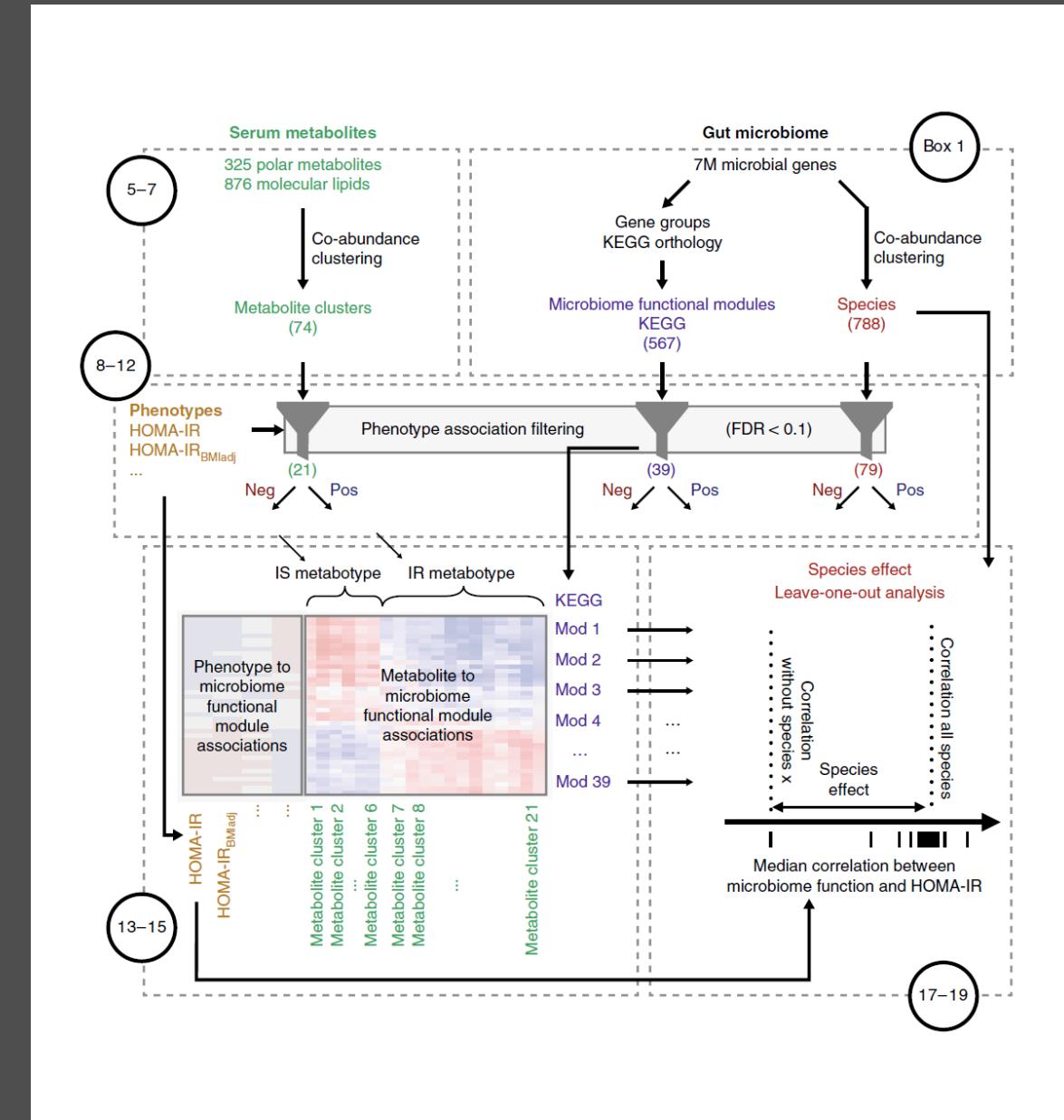
Ridge Regression adjusted for age, sex and BMI

A computational framework to integrate high-throughput '-omics' datasets for the identification of potential mechanistic links

Helle Krogh Pedersen^{1,12}, Sofia K. Forslund^{2,3,12}, Valborg Gudmundsdottir⁴,
Anders Østergaard Petersen⁴, Falk Hildebrand³, Tuulia Hyötyläinen⁵, Trine Nielsen^{ID 1},
Torben Hansen¹, Peer Bork^{ID 3}, S. Dusko Ehrlich^{6,7}, Søren Brunak^{4,8}, Matej Oresic^{9,10},
Oluf Pedersen^{1*}, Henrik Bjørn Nielsen¹¹

A computational framework to integrate high-throughput '-omics' datasets for the identification of potential mechanistic links

Helle Krogh Pedersen^{1,12}, Sofie K. Forslund^{2,3,12}, Valborg Gudmundsdottir⁴,
Anders Østergaard Petersen⁴, Falk Hildebrand³, Tuulia Hyötyläinen⁵, Trine Nielsen^{1,12},
Torben Hansen¹, Peer Bork^{1,3}, S. Dusko Ehrlich^{6,7}, Søren Brunak^{4,8}, Matej Oresic^{9,10},
Oluf Pedersen^{1*}, Henrik Bjørn Nielsen¹¹



A computational framework to integrate high-throughput ‘-omics’ datasets for the identification of potential mechanistic links

Helle Krogh Pedersen^{1,2}, Sofia K. Forslund^{2,3,12}, Valborg Gudmundsdottir⁴,
Anders Østergaard Petersen⁴, Falk Hildebrand⁴, Tuulia Hyötyläinen⁵, Trine Nielsen^{6,7},
Torben Hansen¹, Peer Bork^{8,9}, S. Dusko Ehrlich^{6,7}, Søren Brunak^{4,8}, Matej Oresic^{9,10},
Oluif Pedersen¹¹, Henrik Bjørn Nielsen¹¹

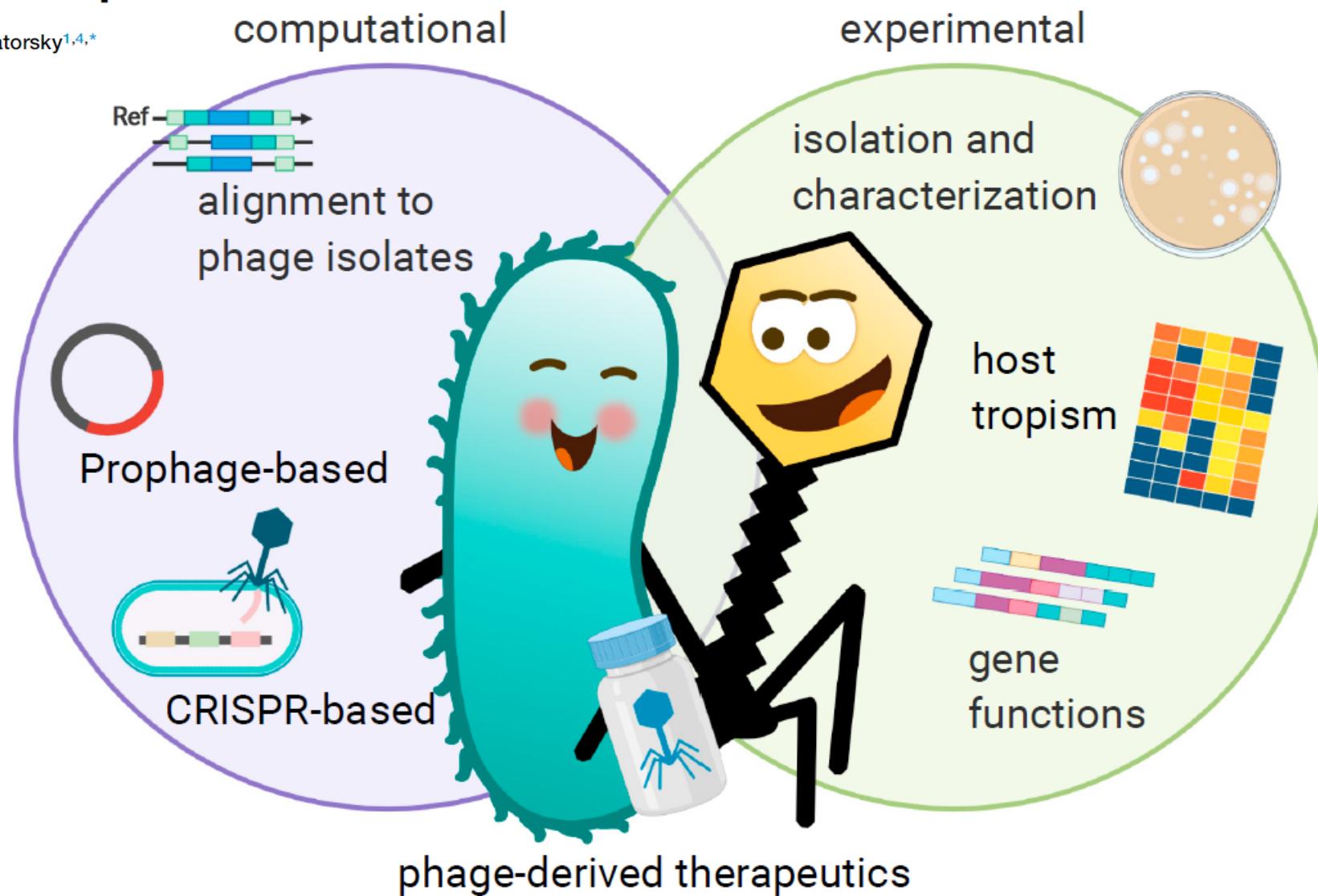
Table 1 | Examples of different strategies for data-driven dimensionality reduction and resources that can be applied for knowledge-driven dimensionality reduction

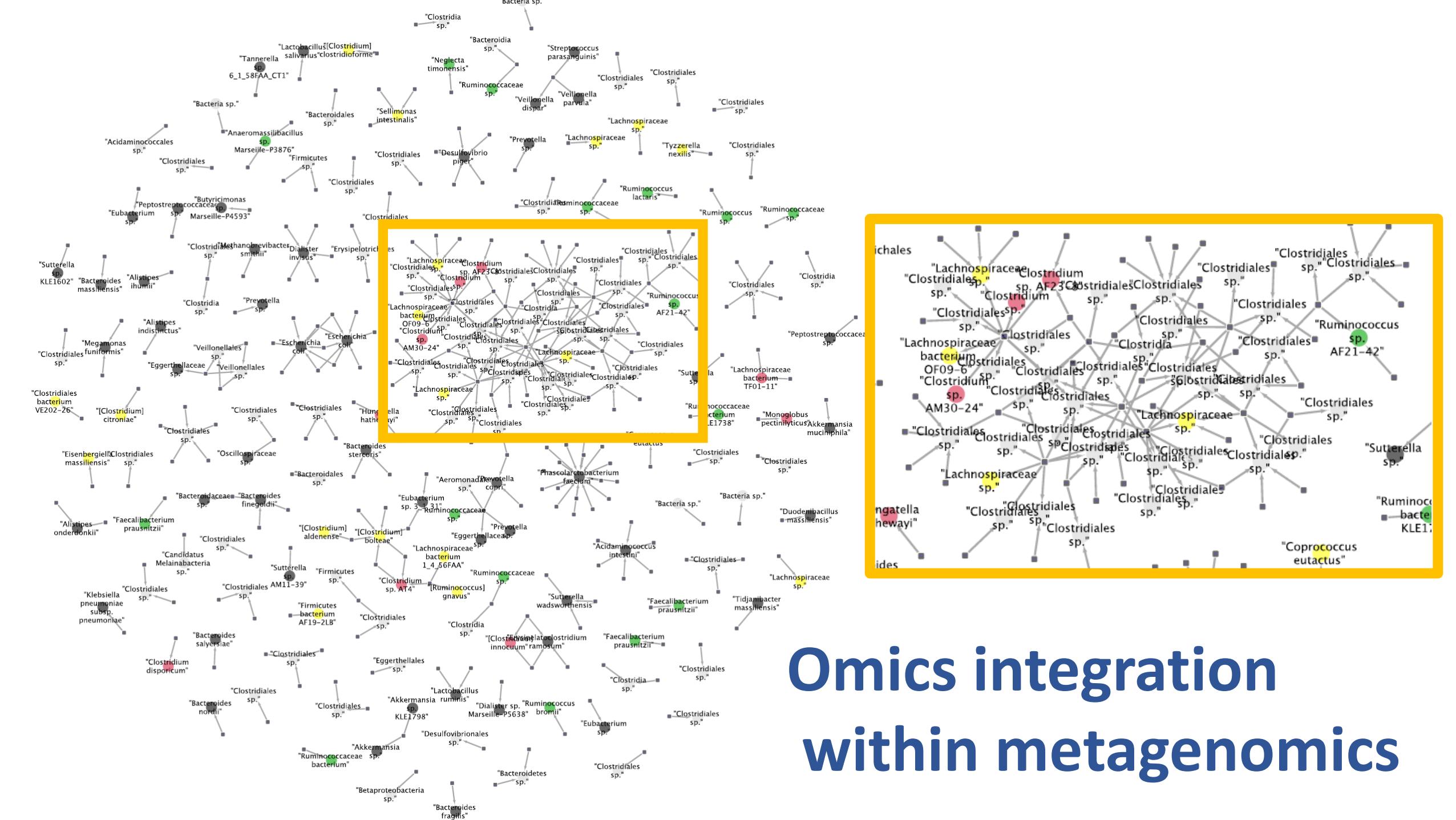
	Data driven	Knowledge driven
Microbiome data	Binning of co-abundant genes, e.g., using the MGS framework ^a ¹² , MetaBat ⁵⁰ , MaxBin ⁵¹ and so on, or using single-copy phylogenetic marker genes (mOTU) ⁵² , or including base composition information using, e.g., CONCOCT ⁵³ . Reference genome-based methods, e.g., MetaPhlAn ⁵⁴	KEGG pathways ^a ⁵⁵⁻⁵⁷ MetaCyc ⁵⁸ Clusters of orthologous groups (COGs) ⁵⁹ Carbohydrate-active enzyme (CAZy) families (http://www.cazy.org) ⁶⁰ Gut metabolic modules (GMMs) ⁶¹
Metabolome data	Clustering of co-abundant metabolites, e.g., using the WGCNA framework ^a ²⁶ or any other unsupervised clustering framework Principal component analysis (PCA) ⁶² Non-negative matrix factorization ⁶³ (NMF)	KEGG pathways ⁵⁵⁻⁵⁷ ConsensusPathDB ⁶⁴ (collection of pathways and metabolite sets from multiple databases, including KEGG)

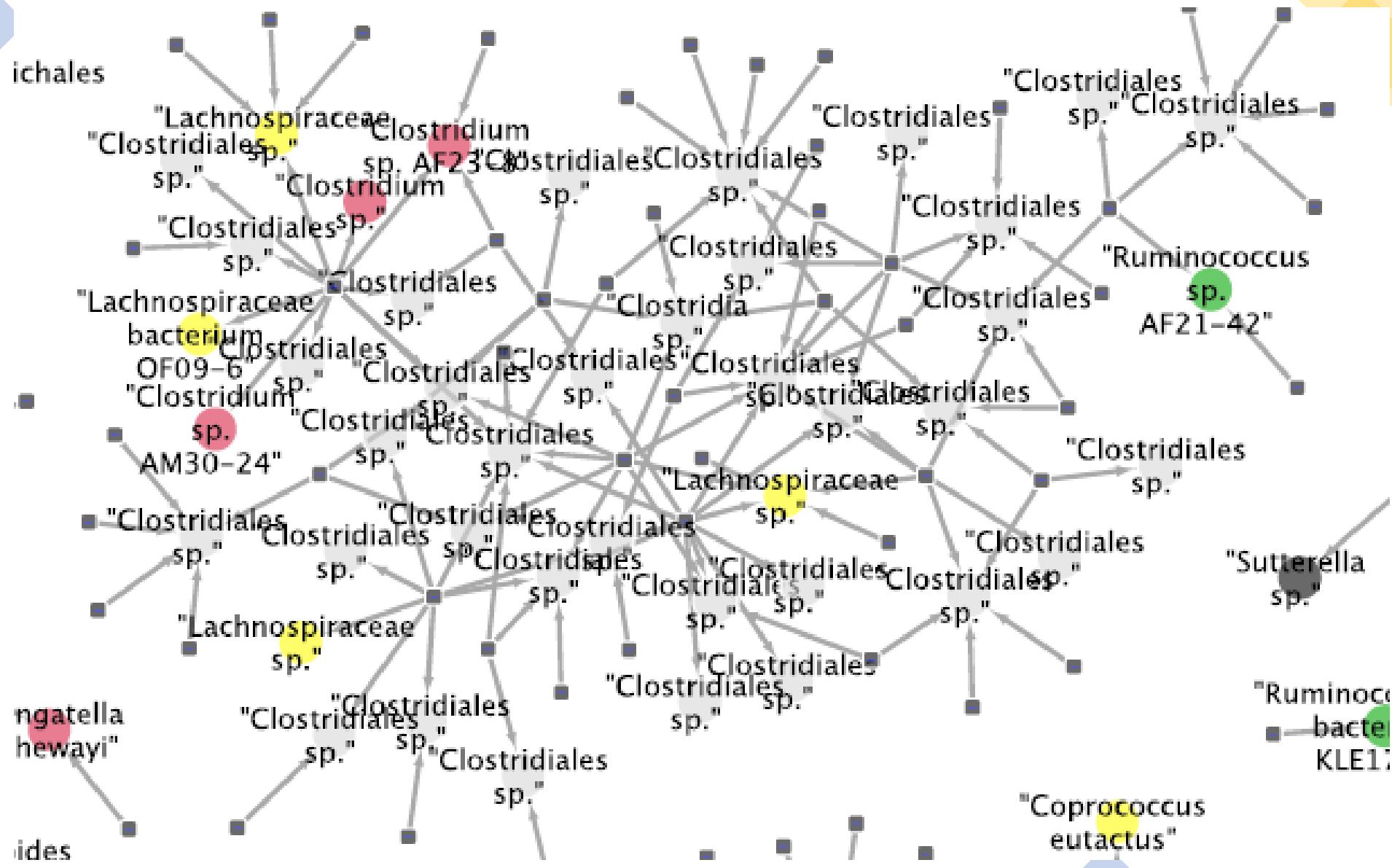
^aIndicates strategies and resources used in this protocol.

Phage-Bacteria Associations: Analyze. Match. Develop Therapies.

Shaqed Carasso,¹ Haitham Hajjo,^{1,2,3} and Naama Geva-Zatorsky^{1,4,*}







**Have you understood
what kind of time we are living
in multi-omics research?**