Propagation of chemical families from high-confidence level metabolite identification through molecular networking in the context of microbiome research

Romina PACHECO TAPIA[1,2], Francesc PUIG CASTELLVI[1,2], Inés CASTRO[1], Maxime Deslande[1], Amélie Bonnefond[1], Philippe Froguel[1,4], Marc-Emmanuel DUMAS[1,2,3,5]

[1]INSERM U1283, CNRS UMR 8199, Institut Pasteur de Lille, University of Lille, Lille University Hospital, Lille, France; [2] Section of Biomolecular Medicine, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK; [3]Section of Genomic & Environmental Medicine, National Heart & Lung Institute, Imperial College London; [4]Department of Metabolism, Section of Genetics and Genomics, Imperial College London, London, UK; [5]McGill Genome Centre, McGill University, Montréal, Qc, Canada.

BACKGROUND

Host-associated samples subjected to untargeted metabolomics have provided valuable insights into how microbes influence health in a bidirectional way[1,2]. However, accurate metabolite annotation and identification remain to be a challenge[3] along with ensuring analytical reproducibility and feature coverage for large cohorts of data[4].

METHODOLOGY

Sample prep

Sigma – Aldrich® Human serum

Sample extraction

Sample reconstitution

LC methods

Optimized workflow

Thermo Scientific™ Hypersil Gold™ C18
Reverse phase (apolar)

Waters ACQUITY BEH Amide
Hydrophilic interaction (polar)

AcquireX Deep-Scan

Data-dependent acquisition & Intelligent data acquisition

ESI +/-

Thermo Scientific™ Vanquish™ UHPLC system and Orbitrap Exploris 240 Mass spectrometer

Data processing tools

Features filtered by:
- Linearity
- Reproducibility
- Repeatability

Feature annotation

IROA TECHNOLOGIES ...making metabolic profiling easy®

★ LEVEL 1

In house chemical library

Data matrix

Harmonization workflow

GNPS

★ LEVEL 2
- INCHIKEY
- IUPAC names
- Automated chemical classification ClassyFire

Molecular networking

- **Featured-based Molecular Networking**
  - Ion Identity Molecular Networking
- **Network Annotation Propagation (NAP)**
  - MetFrag fragmentation *in silico*
  - **MolNetEnhancer**

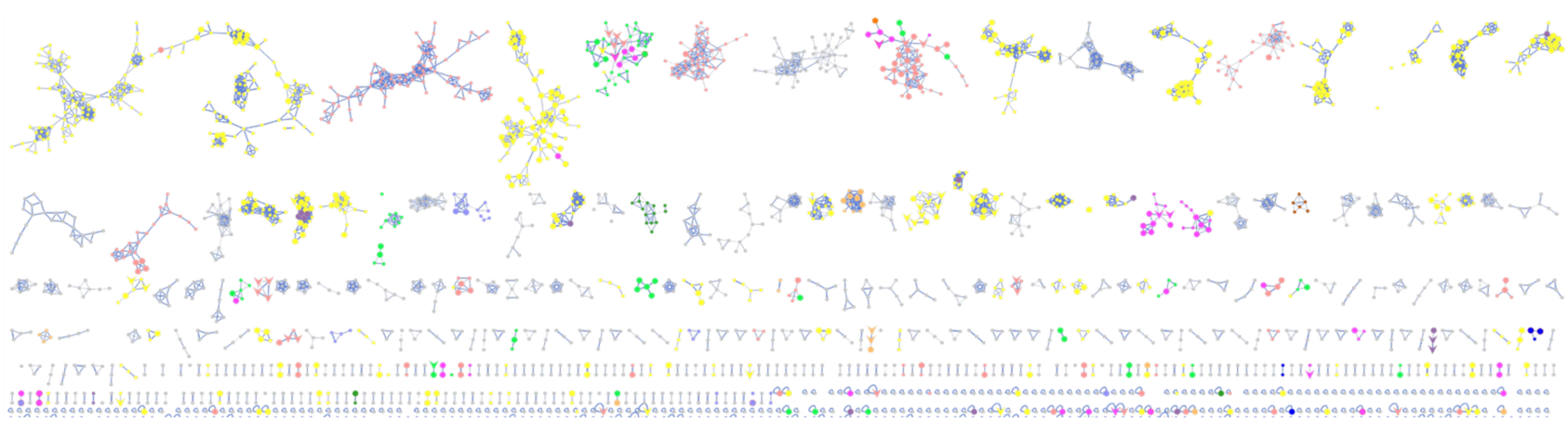
MGF file : Precursor mass, charge and *m/z* - rt pairs

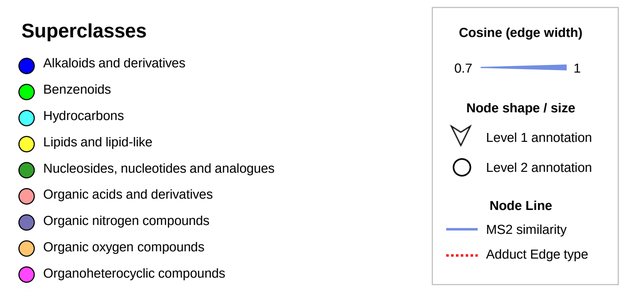
Figure 1. Methods from sample processing to accurate metabolite annotation and propagation. Human serum samples were analyzed using a UHPLC Vanquish Duo coupled to a high resolution Orbitrap Exploris™ 240 mass spectrometer with two optimized methods for polar and non-polar metabolites in negative and positive electrospray ionization mode. Intelligent data acquisition workflow was implemented in addition to the Data-Dependent Acquisition method in order to increase spectral data required for metabolite annotation. An ion Identity Molecular networking (IIMN) strategy was applied using the GNPS on-line platform to expand the chemical class starting from the known metabolites, annotated with both public spectral reference library (GNPS) and an in-house spectral library, to the unknowns.

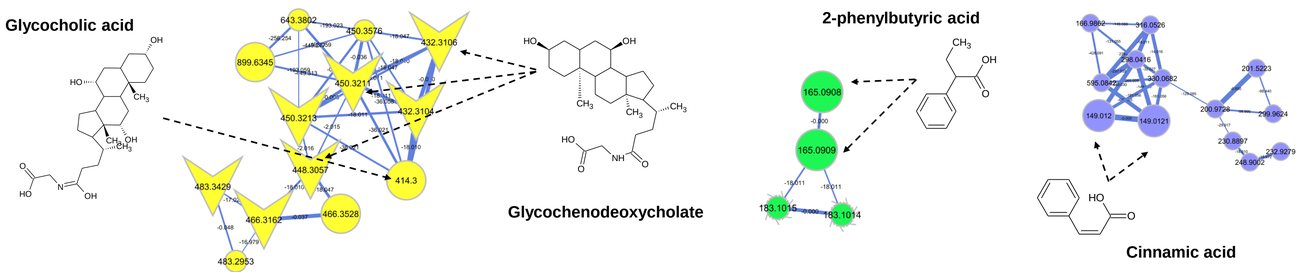
RESULTS

The metabolome profiling of the samples provided a total of 4840 linear and reproducible features (*m/z*-rt pairs) detected in positive and negative mode with both LC-MS methods. Focusing on the positive ESI mode, 17% of the detected features were annotated and 15% of which correspond to a high confidence level annotation. Among them, we found lipids (47%), organic acids and derivatives (20%), organoheterocyclic compounds (13%), benzenoids (9%), organic oxygen compounds (3%) and other chemical superclasses, 49 classes and over 80 subclasses (**Figure 2**). IIMN also allowed us to reduce redundancies of ion species and to expand the chemical information of the unannotated metabolites (**Figure 3**).

Figure 2. Collapse Ion Identity Molecular Networking from features in positive ESI mode with expanded superclasses

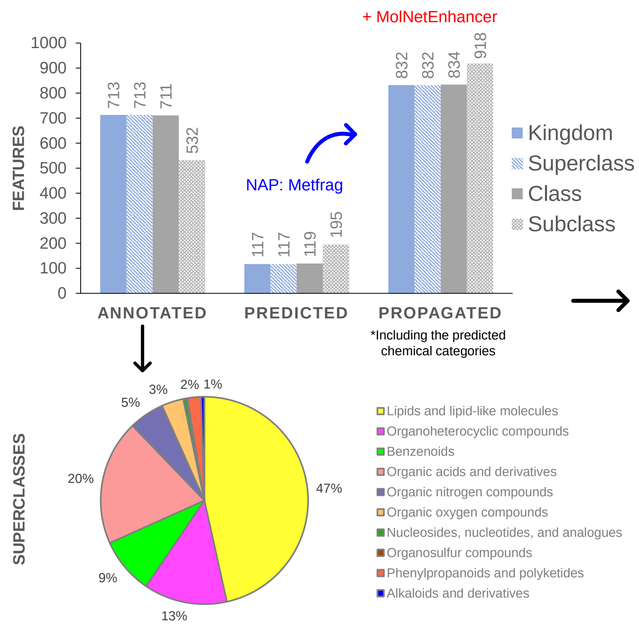


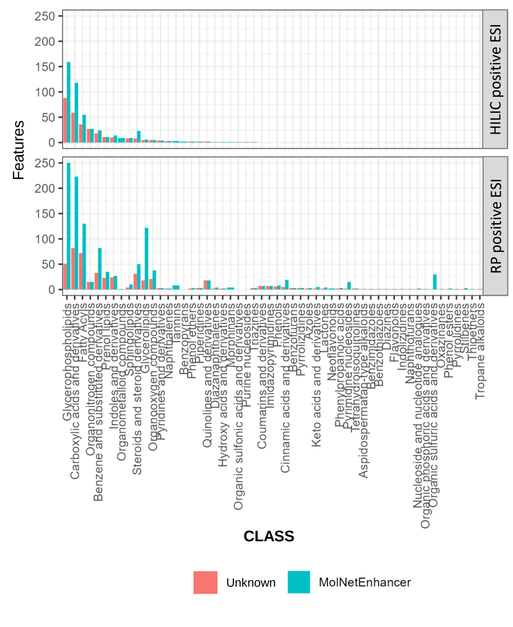




- ✓ *In silico* fragmentation tools (MetFrag) and MolNetEnhancer allowed an increase of 20% in the chemical categories assignment of the features detected in positive ESI mode.
- ✓ This combination of approaches represent a novel strategy in untargeted metabolomics applying a high-confidence level of annotation.

Figure. 3 Propagation of chemical classification with MolNetEnhancer from features in positive ESI mode





CONCLUSION

We have applied *in silico* approaches[5] for metabolite annotation as it harness advanced machine learning and predict fragmentation spectra from known structures. This will be essential for the implementation of a reproducible workflow for untargeted LCMS analysis of biofluids in the context of metabolomics in microbiome research. It will also help to increase and improve the identification of metabolites of interest to provide an appropriate biological interpretation.

ACKNOWLEGDEMENTS

This research was conducted within the context of the CNRS–Imperial International Research Project METABO-LIC and was supported by EGID (ANR-10-LABX-6), Precidiab (ANR-18-IBHU-0001) and the "Accueil de Talent" grant joint awarded by site ULNE (ANR-16-IDEX-0004-ULNE,R-002-20-TALENT-DUMAS), the Hauts-de-France Regional Council (20002845, 20001891 / NP0025517) and the Métropôle Européenne de Lille (MEL,2019_ESR_11).

BIBLIOGRAPHY

1. Bauermeister, A., Mannochio-Russo, H., Costa-Lotufo, L.V., Jarmusch, A.K., Dorrestein, P.C., 2022. Nat. Rev. Microbiol. 20, 143–160.
2. Dekkers, K.F., Sayols-Baixeras, S., Baldanzi, G. et al., 2022. Nat Commun 13, 5370.
3 Plumb, R.S., Gethings, L.A., Rainville, P.D., Isaac, G., Trengove, R., King, A.M., Wilson, I.D., 2023. TrAC Trends Anal. Chem. 160, 116954.
4. Zhou, Z., Luo, M., Zhang, H., Yin, Y., Cai, Y., Zhu, Z.-J., 2022. Nat. Commun. 13, 6656.
5. Ernst, M.; Kang, K.B.; Caraballo-Rodríguez, A.M.; Nothias, L.-F.; Wandy, J.; Chen, C.; Wang, M.; Rogers, S.; Medema, M.H.; Dorrestein, P.C.; et al. 2019. Metabolites. 9, 144.