Reverse Metabolomics for the Discovery of Chemical Structures from Humans

Chua Huiyi Sandy, Foo Chuan Jen Jonathan, Abel Chen Kai Yi

INTRODUCTION

The field of metabolomics is rapidly expanding in scope and depth. However, there remains challenges in determining the structures and phenotypic contexts of untargeted metabolomics experiments. It should also be noted that the annotation of the human metabolome is incomplete and current methodologies are lacking in efficiency and throughput. In a typical untargeted metabolomics study involving human-derived samples, only about 10% of the data can be annotated with a known structure[1-5].

As a result, new strategies are necessary to identify and characterize the molecules present in humans. Understanding their structures is essential for hypothesizing their synthesis, roles in biochemical pathways, and potential contributions to disease development and progression.

One such disease is Inflammatory Bowel Disease (IBD), which is characterised by a chronic inflammation of the gastrointestinal tract and encompasses Crohn's Disease (CD) and Ulcerative Colitis (UC). Currently, there is a lack of a comprehensive molecular understanding of its pathogenesis, hindering advances in its diagnosis and treatment.

In this paper, **a novel discovery strategy of reverse metabolomics is presented**. Tandem mass spectrometry (MS/MS) data from newly synthesized compounds are analyzed against public metabolomics databases to identify potential phenotypic correlations. This uncovered conjugated bile acids that display strong associations with IBD, providing an inlet to discover potential pathways contributing to its pathogenesis, creating new avenues for diagnosis and management.

There are three main phases of this study:

Phase 1: Broad synthesis and exploration of metabolite classes (N-acyl amides, fatty acid esters, bild acid esters and conjugated bile acids) were conducted
Phase 2: Repository-scale analysis which facilitated the discovery of conjugated bile acids associated with IBD
Phase 3: Validation of the above analysis using human IBD cohorts

METHODS

Phase 1: Conducting broad synthesis and exploration of metabolite classes

Fig. 1a | Workflow for the reverse metabolomics strategy using LC-MS/MS and MASST and ReDU data analysis tools and platforms.

Liquid chromatography-mass spectrometry (LC-MS/MS) spectra were created for four metabolite classes of interests (N-acyl amides, fatty acid esters, bile acid esters and conjugated bile acids). They were then searched for in public metabolomic datasets[6] using the mass spectrometry search tool (MAAST)[7], and the sample information from the available reanalysis of data user interface (ReDU)[8] was summarised. This helped to to uncover phenotypic associations (Fig. 1a).

In particular, the synthesis of acyl amides and esters, as well as bile acid conjugates are as follows:

Synthesis of Acyl Amides and Esters

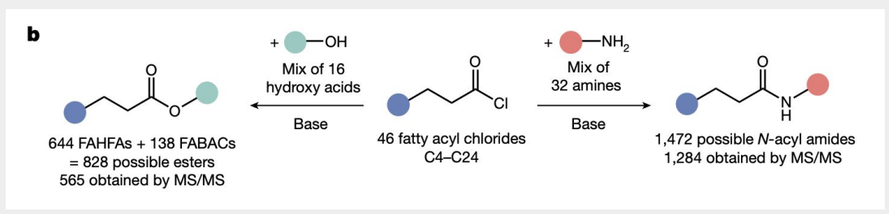

Fig. 1b | Synthesis scheme for acyl amides and esters

A library of acyl amides and esters was synthesized through reactions of 46 fatty acyl chlorides with 32 amines and 17 hydroxy acids. Subjecting them to a search in public metabolomic datasets yielded 1472 acyl amides and 782 esters (Fig. 1b).

Synthesis of Bile Acid Conjugates

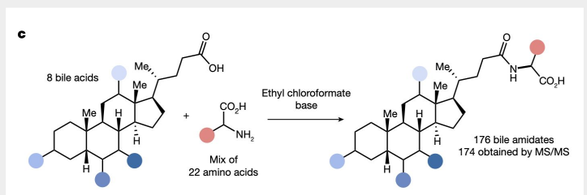

Fig. 1c | Combinatorial bile acid conjugation reaction performed for the discovery of bile acid amidates

Combinatorial amide coupling reactions were carried out between 8 dihydroxylated and trihydroxylated bile acids and 22 amino acids. MS/MS data was then collected for each combination of bile acid and amino acid pair (Fig.1c), with matches found to 145 of the synthesised compounds in public metabolomics data.

Phase 2: Discovery of conjugated bile acids that are associated with inflammatory bowel disease (IBD)

In the second part of the experiment, the team examined the association of different conjugated bile acid structures with health status in public data. It was found that certain conjugations were observed more frequently in all types of IBD samples.

Phase 3: Validation of the above analysis using human IBD cohorts

The team set out to confirm these findings, starting with the collection of data and samples from the longitudinal integrative Human Microbiome Project 2 (iHMP2) and cross-sectional PRISM IBD cohorts. Fecal samples were collected from participants in the following groups: 1. Without IBD 2. With UC 3. With CD. Fecal extracts were then pooled and analysed with the mixed bile acid standards to match retention times and MS/MS spectra.

RESULTS

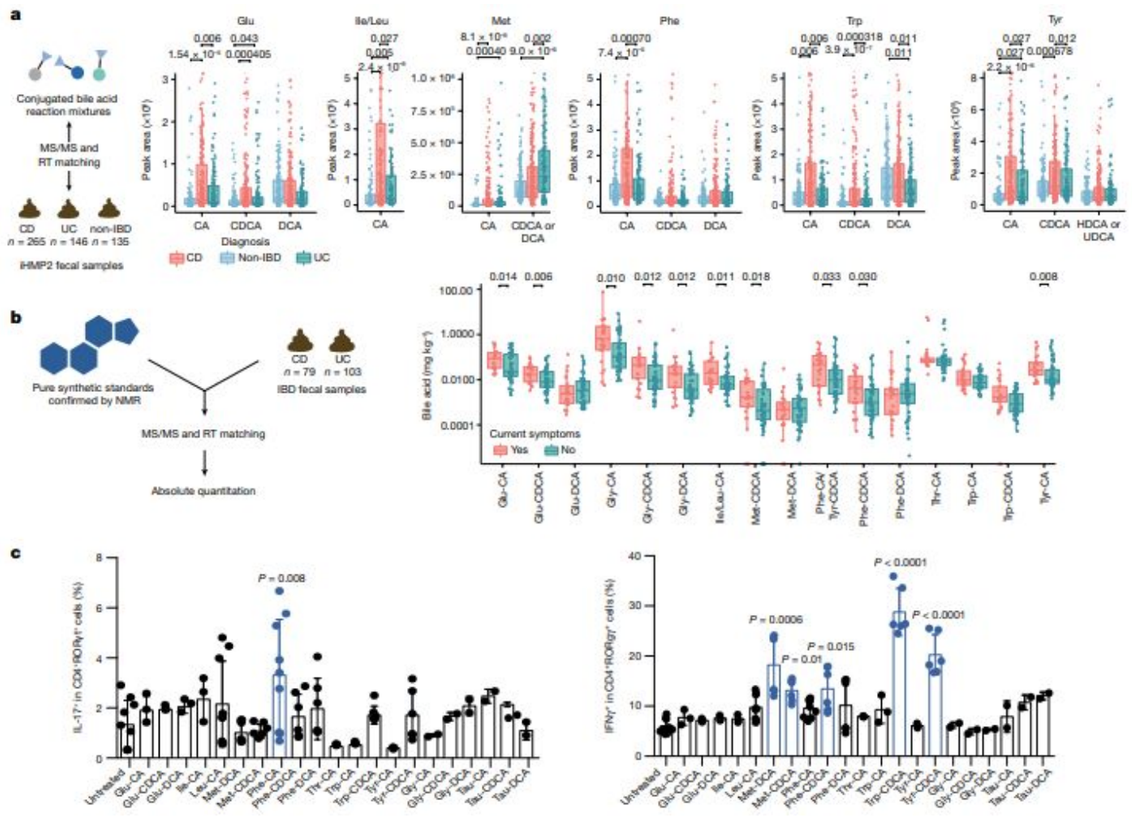

Fig. 2 | IBD association of new conjugated bile acids. a, Examples of retention time (RT) matching and MS/MS spectra of a standard to a pooled fraction for which only one isomer matched by retention time and for which isomers could not be resolved. Relative peak area abundances of selected bile acids that were higher in patients CD (red) and/or in patients with UC (green) compared with individuals without IBD (blue) in the iHMP2 study. Boxplots show first (lower) quartile, median and third (upper) quartile with whiskers as 1.5 times the interquartile range. CD, n = 265; non-IBD, n = 135; UC, n = 146. **b**, Concentrations of conjugated bile acids in fecal samples from individuals with active (n = 23) or inactive (n = 72) Crohn's disease. Values on the y axis represent mg of bile acid per kg of fecal matter. **c**, Flow cytometric quantification of IL-17 (left) and IFNγ (right) production in naive CD4+ T cells from Foxp3-hCD2 reporter mice. Cells were treated with 100 μM of bile acids on day 0 and CD4+ T cells were gated for analyses on day 3. n = 6 for controls and n > 3 for biologically independent samples for every substrate tested.

- **Bile acids conjugated to the amino acids citrulline (Cit), glutamic acid (Glu), histidine (His), isoleucine/leucine (Ile/Leu), phenylalanine (Phe), threonine (Thr), tryptophan (Trp) and tyrosine (Tyr) were detected in higher abundance in CD patients** compared to individuals without IBD (Fig. 2a)

- Specifically, **primary bile acid amidates were higher in abundance in CD**, whereas amidates of deoxycholates and related secondary bile acid isomers remain unchanged relative to samples from individuals without IBD

- When examining how the abundance of conjugated bile acids relate to symptom activity in CD compared to UC, it was found that **11 out of 19 bile acids that were quantified were significantly increased in individuals with active symptoms**, but only in the CD group (Fig. 2b)

- Five of the synthesised conjugated bile acids (Met-CDCA, Met-DCA, Phe-CDCA, Trp-CDCA and Tyr-CDCA) had elicited **increased levels of interferon-γ (IFNγ)** (Fig. 2c), a key cytokine involved in the pathogenesis of CD[9]

- **Thr-CA, Glu-DCA and Glu-CDCA were found to act as pregnane X receptor (PXR) agonists** involved in xenobiotic transport and metabolism. Decreased PXR expression has been associated with IBD and CD[10]

Taken together, these findings support the hypothesis that some of the newly discovered bile acids play important roles in IBD through PXR and/or immune-related processes, underscoring the utility of specific bile acids as biomarkers for CD and their relevance in IBD research and diagnostics. The results could also suggest why some patients are unresponsive to rifaximin (PXR agonist) treatment in clinical studies, as they may already possess large quantities of PXR agonists produced by their microbiota.

DISCUSSION & CONCLUSION

Yield and Versatility of Reverse Metabolomics
The reverse metabolomics approach resulted in an expansion of metabolic knowledge using simple chemical transformations. Notably, **relevant disease associations were found from a narrow range of 3 metabolite classes which yielded 2000 relevant unique compounds**. This method can be expanded to any synthetically accessible compound class that can ionize and fragment in a mass spectrometer.

Potential for Disease Phenotyping and Metabolomic Matching
The results demonstrated that **disease phenotypes could be connected to newly discovered bile acids, something that may eventually have diagnostic value to accurately and non-invasively detect CD**. It was also found that these bile acids may mediate IBD processes through the regulation of host immune function and PXR signalling. These new discoveries made on IBD sheds light on the ability of reverse metabolomics to advance our understanding, diagnosis and management of disease, and could potentially be utilised for an even greater number of conditions, spelling exciting possibilities for the future of healthcare.

Limitations regarding Reverse Metabolomics
Firstly, metabolites must be present in sufficient quantities or ionize well to be detected and selected for MS/MS acquisition, as reverse metabolomics requires a MS/MS spectrum. This limitation was addressed through combinatorial synthesis. This strategy, however, requires the investigator to have some hypothesis with respect to structures of interest. *Secondly*, isolation and NMR/X-ray structural analyses was not conducted, and therefore, there is a possibility that different isomers are represented by the data. *Lastly*, the 1.2 billion spectra that are currently part of MASST searches do not encompass all molecules. This limits re-analyses of public data in the discovery phase of this approach.

REFERENCES

Original Paper: Gentry, E.C., Collins, S.L., Panitchpakdi, M. et al. Reverse metabolomics for the discovery of chemical structures from humans. Nature 626, 419–426 (2024). https://doi.org/10.1038/s41586-023-06906-8

1. Aksenov, A. A., da Silva, R., Knight, R., Lopes, N. P. & Dorrestein, P. C. Global chemical analysis of biology by mass spectrometry. Nat. Rev. Chem. 1, 0054 (2017).

2. Blaženović, I. et al. Structure annotation of all mass spectra in untargeted metabolomics. Anal. Chem. 91, 2155–2162 (2019).

3. Hassanpour, N. et al. Biological filtering and substrate promiscuity prediction for annotating untargeted metabolomics. Metabolites 10, 160 (2020).

4. Schmid, R. et al. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. Nat. Commun. 12, 3832 (2021).

5. Viant, M. R., Kurland, I. J., Jones, M. R. & Dunn, W. B. How close are we to complete annotation of metabolomes? Curr. Opin. Chem. Biol. 36, 64–69 (2017).

6. Wang, M. et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. 34, 828–837 (2016).

7. Wang, M. et al. Mass spectrometry searches using MASST. Nat. Biotechnol. 38, 23–26 (2020).

8. Jarmusch, A. K. et al. ReDU: a framework to find and reanalyze public mass spectrometry data. Nat. Methods 17, 901–904 (2020).

9. Rovedatti, L. et al. Differential regulation of interleukin 17 and interferon γ production in inflammatory bowel disease. Gut 58, 1629–1636 (2009).

10. Wilson, A., Almousa, A., Teft, W. A. & Kim, R. B. Attenuation of bile acid-mediated FXR and PXR activation in patients with Crohn's disease. Sci. Rep. 10, 1866 (2020).