Integrated systems biology to study non-alcoholic fatty liver disease in obese women

Lesley Hoyles[1]*, José-Manuel Fernández-Real[2]*, Massimo Federici[3]*, Matteo Serino[4,5], Vincent Azalbert[4,5], Vincent Blasco[4,5], James Abbott[1], Richard H. Barton[1], Josep Puig[2], Gemma Xifra[2], Wifredo Ricart[2], Mark Woodbridge[1], Christopher Tomlinson[1], Marina Cardellini[3], Francesca Davato[3], Iris Cardolini[3], Ottavia Porzio[6], Paolo Gentilieschi[6], Frédéric Lopez[4,5], Fabienne Foufelle[7], Catherine Postic[8], Sarah A. Butcher[1], Elaine Holmes[1], Jeremy K. Nicholson[1], Rémy Burcelin[4,5], Marc-Emmanuel Dumas[1]

[1]Surgery and Cancer, MRC-PHE Centre for Environment and Health, Imperial College London; [2]Department of Endocrinology, Diabetes and Nutrition, Universitat of Girona, Spain; [3]Department of Systems Medicine, University of Rome Tor Vergata, Italy; [4]Institut National de la Santé et de la Recherche Médicale (INSERM), Toulouse, France; [5]Université Paul Sabatier (UPS), Institut des Maladies Métaboliques et Cardiovasculaires (I2MC), France; [6]Department of Experimental Medicine and Surgery, University of Rome Tor Vergata, Rome; [7]Sorbonne Universités, UPMC Univ Paris 06, UMRS 1138, Centre de Recherche des Cordeliers, Paris, France

Background

Non-alcoholic fatty liver disease (NAFLD) is a multi-factorial condition and one of the most common causes of chronic liver disease, with its prevalence increasing worldwide as a result of the obesity epidemic[1,2]. Data are available from rodent models regarding the role of the gut microbiota and microbiome in liver disease and their contribution to NAFLD phenome stratification (i.e. a comprehensive set of molecular phenotypes useful to identify subgroups of patients). Microbial factors associated with NAFLD include bacterial lipopolysaccharides (LPS) and bile acid–FXR activators[2,3]. The relevance of such factors in humans remains poorly understood. In addition, gut microbial populations of humans with NAFLD are inadequately characterized, and it is not known if changes in these populations and/or their functions contribute to initiation of NAFLD and/or its progression. We used an integrative multi-omics approach combining shotgun metagenomics (faecal microbiome) and molecular phenomics (liver transcriptome, plasma and urine metabonomes, clinical phenotyping) to decipher multi-scalar interactions in NAFLD in obese women.

Objective

To integrate metagenomic, transcriptomic, metabonomic and clinical data to evaluate the contribution of the faecal microbiome to the molecular phenome (hepatic transcriptome, plasma and urine metabonomes) of NAFLD independent of clinical confounders in morbidly obese women recruited to the FLORINASH study.

Methods

Faecal, liver biopsy, blood and urine samples and data for 28 clinical variables were collected for 56 obese [body mass index (BMI) >35 kg/m^2] women from Italy (n = 31) and Spain (n = 25) who elected for bariatric surgery. Confounder analyses of clinical data were done using linear modeling. Histological examination of liver biopsies was used to grade NAFLD (NAFLD activity score: 0, 1, 2, 3). Faecal metagenomes were generated and analysed using the Imperial Metagenomics Pipeline[4]. Differentially expressed genes were identified in hepatic transcriptomes using limma[5], and analysed using Enrichr[6], network analyses and Signaling Pathway Impact Analysis (SPIA)[7]. ^{1}H-NMR data were generated for plasma and urinary metabonomes[8]. Clinical, metagenomic, transcriptomic and metabonomic data were integrated using partial Spearman's correlation, taking confounders (age, BMI and cohort) into account. Interactions between the faecal microbiome and the clinical/molecular phenome were quantified, allowing generation of a Receiver Operating Characteristic (ROC) curve that confirmed the diagnostic power of molecular phenomic and metagenomic indices.

Results

1. NAFLD activity score was anti-correlated with microbial gene richness, and correlated with abundance of *Proteobacteria*. Microbial gene richness was correlated with clinical markers of NAFLD (Fig. 1).

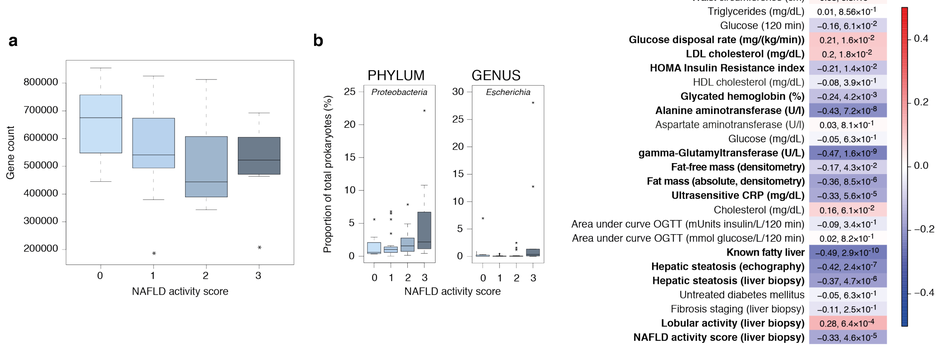


Fig. 1. (a) Association of microbial gene richness and (b) abundance of *Proteobacteria* in the faecal microbiota with NAFLD activity score. (c) Correlation of microbial gene richness with clinical data. *P* values adjusted for multiple correction testing (Benjamini–Hochberg).

2. KEGG analyses of metagenomic data suggested increased microbial processing of dietary lipids and amino acids, as well as endotoxin-related (LPS) processes related to *Proteobacteria* (Fig. 2).

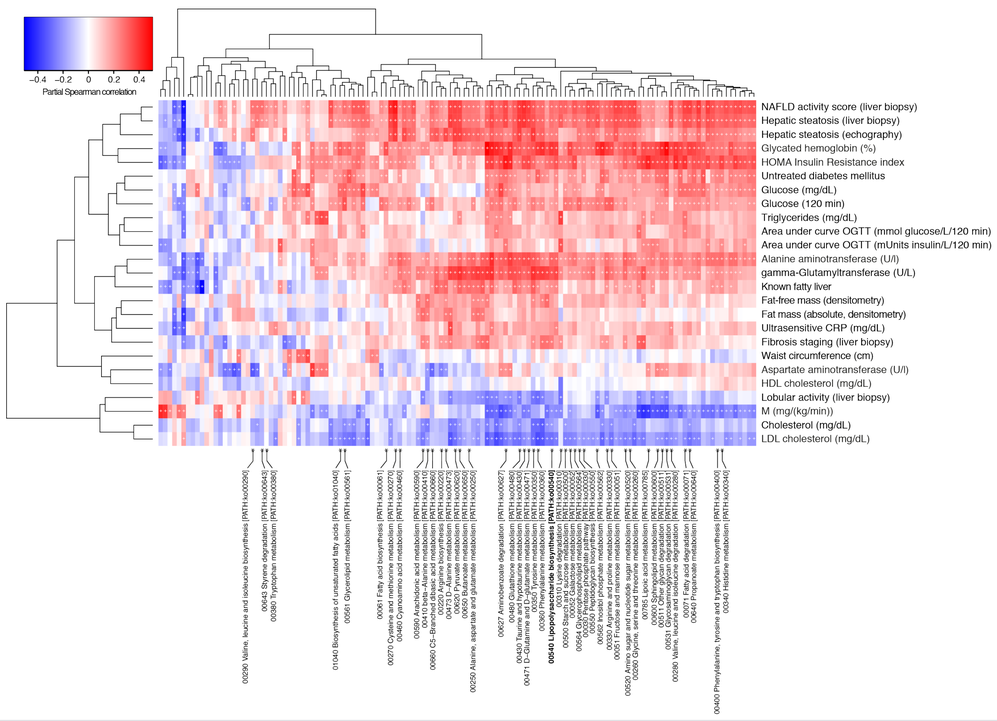


Fig. 2. Heat map showing associations of metagenome-derived KEGG pathway data with clinical data. +, Statistically significant association (P < 0.05, Benjamini–Hochberg).

3. Metabonomic profiles highlighted imbalances in branched-chain amino acid metabolism, gut-derived microbial metabolites resulting from metabolism of amino acids and choline metabolism (Fig. 3).

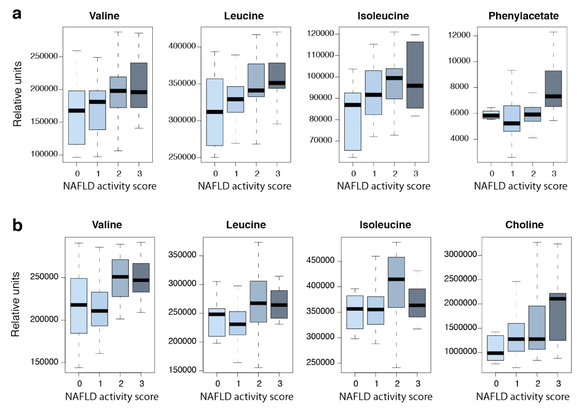
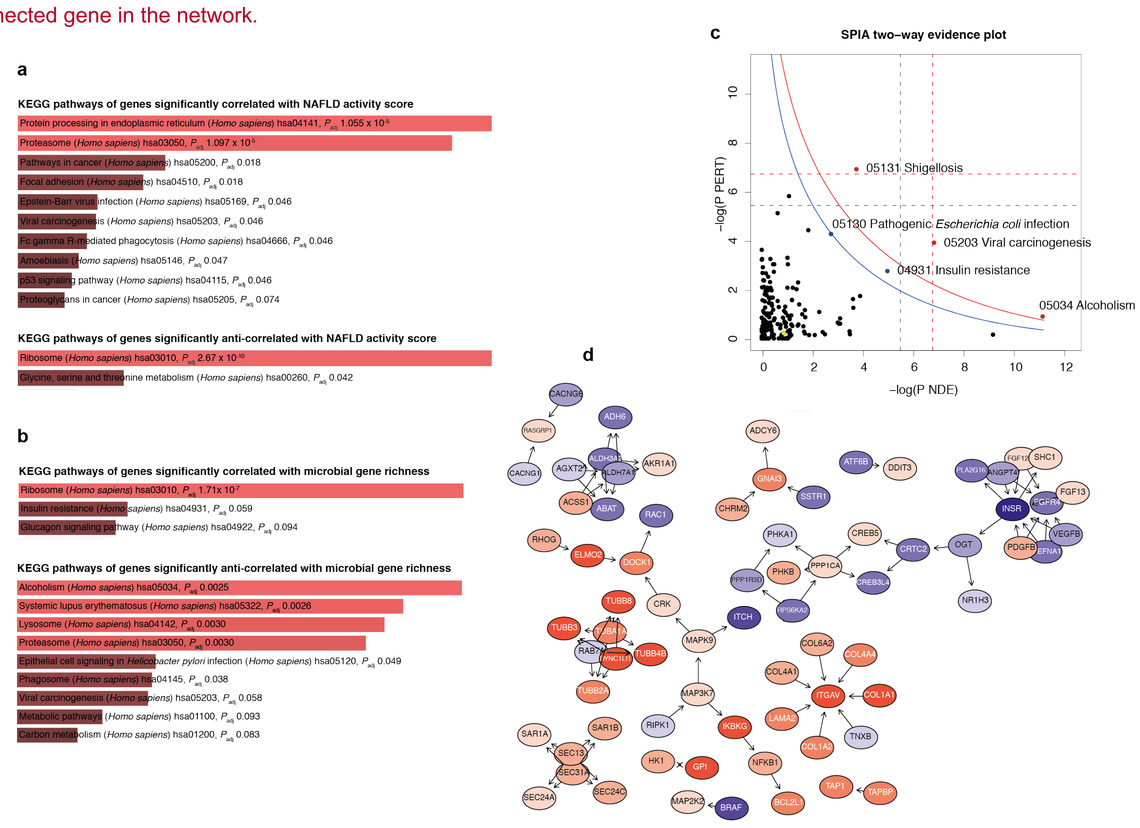


Fig. 3. (a) Plasma metabolites most significantly (P < 0.05, Benjamini–Hochberg) partially correlated with NAFLD activity score. First three panels, branched-chain amino acids; phenylacetate is derived from microbial use of phenylalanine. (b) Urinary metabolites most significantly (P < 0.05, Benjamini–Hochberg) partially correlated with NAFLD activity score. First three panels, branched-chain amino acids. Our data suggest choline is excreted rather than used by the gut microbiota.

4. NAFLD-associated hepatic transcriptomes were associated with branched-chain amino acid metabolism, endoplasmic reticulum/phagosome. Hepatic genes significantly correlated with NAFLD activity score and microbial gene richness were significantly associated with immune responses associated with non-specific microbial infections and insulin resistance, and the most connected gene was *INSR* (insulin receptor) (Fig. 4).

Fig. 4. KEGG pathways over-represented in the genes significantly correlated with (a) NAFLD activity score and (b) microbial gene richness. Only significant results (P < 0.1, Benjamini–Hochberg) are shown. (c) SPIA analysis of NAFLD activity score–microbial gene richness intersecting genes. (d) Network analysis of the NAFLD activity score–microbial gene richness intersecting genes. The correlation values for NAFLD activity score were used to generate the network: the bluer a node, the more significantly anti-correlated NAFLD activity score is with the hepatic gene; the redder a node, the more significantly correlated NAFLD activity score is with the hepatic gene. Analysis of betweenness centrality showed *INSR* to be the most connected gene in the network.



5. Molecular phenomic signatures were stable and predictive regardless of sample size, and consistent with the microbiome making a significant contribution to the NAFLD phenome (Fig. 5).

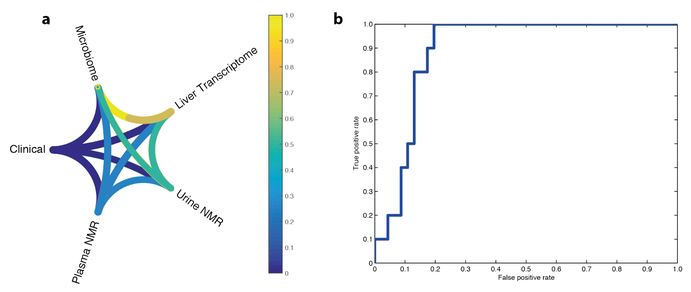


Fig. 5. Phenome-wide crosstalk and predictive analyses. (a) Metagenome–phenome matrix correlation network computed for the patients using the modified Rv correlation matrix coefficient. (b) Performance of classification of NAFLD activity score 3 status (n = 9, vs others, n = 47) based on matching molecular phenome and faecal metagenome profiles. A Receiver-Operator Characteristic (ROC) curve was obtained for the cross-validated model predictions derived from an O-PLS-DA model, reaching an AUC of 0.8913, corresponding to the successful prediction rate.

Conclusions

i) Low microbial gene richness is associated with NAFLD.

ii) There is disruption of the gut–liver axis in NAFLD, which can be seen in the faecal microbiome, hepatic transcriptome and urinary and plasma metabonomes.

iii) Consistency of phenome signatures strongly supports a relationship between microbial amino acid metabolism and microbial gene richness, hepatic gene expression and biofluid metabonomes, and ultimately NAFLD activity scores.

iv) There is a close association between the faecal microbiome, plasma/urinary metabonomes, hepatic steatosis, and clinical and molecular insulin resistance in morbid obesity.

v) Translational validation of rodent model data demonstrating an interplay between the microbiome and host gene expression in inflammation and host metabolism.

*These authors made equal contributions to the work.

Acknowledgements

This work was supported by EU-FP7 FLORINASH (Health-F2-2009-241913) to RB, MF, JMFR, FF, CP, EH and JKN. This work used the computing resources of the UK MEDical BIOinformatics partnership – aggregation, integration, visualization and analysis of large, complex data (UK MED-BIO), which is supported by the Medical Research Council (grant number MR/L01632X/1). LH is in receipt of an MRC Intermediate Research Fellowship in Data Science (grant number MR/L01632X/1). This work was also supported by funding to RB (Region 2009-2014; Agence Nationale de la Recherche ANR Bactimmunodia & GAD), and to JMFR (CIBERobn Pathophysiology of Obesity and Nutrition and FEDER funds).

Contact
Lesley Hoyles
lesley.hoyles11@imperial.ac.uk

References

[1] Abu-Shanab & Quigley (2010). *Nat Rev Gastroenterol Hepatol* **7**, 691–701.
[2] Dumas *et al.* (2014). *Gastroenterology* **146**, 46–62.
[3] Houle *et al.* (2010). *Nat Rev Genet* **11**, 855–866.
[4] Abbott *et al.* (2017). In preparation.
[5] Ritchie *et al.* (2015). *Nucleic Acids Res* **43**, e47.
[6] Chen *et al.* (2013). *BMC Bioinformatics* **14**, 128.
[7] Tarca *et al.* (2009). *Bioinformatics* **25**, 75–82.
[8] Dumas *et al.* (2005). *Proc Natl Acad Sci USA* **103**, 12511–12516.