Identifying transient and stable bacteria- metabolite interactions from longitudinal multi-omics data

Dan Zhang
West China Second University Hospital of Sichuan University
Benjamin H Mullish
St Mary's Hospital Campus, Imperial College London
Jinfeng Wang
China Agricultural University
Grace Barker
St Mary's Hospital Campus, Imperial College London
Despoina Chrysostomou
St Mary's Hospital Campus, Imperial College London
Shengtao Gao
China Agricultural University
Lu Chen
West China Second University Hospital of Sichuan University
Julie A K McDonald
St Mary's Hospital Campus, Imperial College London
Julian R. Marchesi
St Mary's Hospital Campus, Imperial College London
Lu Cheng (lu.cheng.ac@gmail.com)
Aalto University

Short Report

Keywords: gut microbiome, metabolome, bile acids, interaction, networks, longitudinal

Posted Date: May 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1619373/v1

Abstract

Background

Understanding the complex relationships between bacteria and metabolites in ecological systems are extremely important in studies of different microbiomes. Longitudinal multi-omics study is adopted to investigate interactions between bacteria and metabolites, by directly associating their longitudinal profiles. Since a bacteria/metabolite may involve in many different biological processes, the longitudinal profile is an average of different interactions. Therefore, direct association could only uncover the strongest interactions.

Results

Here we present a computational approach that can rebuild short- and long-term bacteria-metabolite interactions from longitudinal multi-omics datasets. For this task, we re-analyse data (both microbial sequencing and metabolomic analysis) from an *in vitro* model of *Clostridioides difficile* infection and faecal microbiota transplant, a disease state and mode of therapy in which perturbed microbiome-metabolome interactions (and their reversal) are well-established to be pertinent. By analysing such a dataset, we generated both a short-term and a long-term interaction network, which predicted many new interactions. Four new interactions were randomly selected to be validated. In batch culture experiments, we validated two of them: (1) *Ruminococcus gnavus* and *Ruminococcus luti* could generate 3-ketocholanic acid (2) *Blautia obeum* could consume succinate.

Conclusions

The deconvolution of the raw longitudinal signal into short- and long-term trends can help users to gain a deeper understanding of their data. This tool will be useful for high-throughput screening of microbe/metabolite/host interactions from a longitudinal multi-omics setting.

Background

A range of different diseases have been suggested to be associated with distinctive microbiota compositions. Each microbiota is a highly complex ecological environment where host, bacteria and metabolites interplay with each other[1,2]. Bacteria may compete for the same metabolites or cross-feed each other by producing complementary metabolites. Due to the dynamic nature of the micro-ecosystem, longitudinal multi-omics studies are by far the most promising strategy to unravel the underlying interactions.

Conversely, the majority of reported microbiome studies are cross-sectional, which only reflect a snapshot of the whole dynamic interaction process and thus the derived interaction networks are less reliable.

Direct association of subject (bacteria, metabolite) profiles in a longitudinal study[3,4,5,6], however, provides only limited biological insights since an observed measurement is an average of many different processes. Only the strongest interactions can be identified, while weaker interactions are often false positives.

To solve this problem, we used an additive Gaussian process based method, LonGP[7], to deconvolute the original longitudinal signal into short-term and long-term trends. Short-term trends refer to processes that are transient, e.g. swift responses to environmental stimuli, while long-term trends are associated with stable processes due to competitions or symbiosis between bacteria. These decomposed trends reflect the underlying mechanisms from different perspectives, capturing more relevant information than the raw data. Therefore, multiple interaction networks derived from the deconvoluted trends - instead of one derived from raw data - can uncover the underlying biological mechanisms more faithfully. For this task, we performed re-analysis of a previously-reported dataset of longitudinal 16S rRNA gene sequencing and metabolomic data (global metabolite profiling from ^{1}H-NMR, and bile acid profiling from mass spectrometry) from a chemostat (*in vitro*) distal gut model of *Clostridioides difficile* infection (CDI) and faecal microbiota transplant (FMT). A chemostat model dataset is attractive for this purposed as it provides longitudinal data in a well-characterised system, while avoiding confounders found in human studies (e.g. variable diet), while a CDI/FMT dataset is attractive given the well-established contribution of microbiome-metabolome interactions to this condition/ therapeutic intervention[8].

Methods

Data preprocessing

Preprocessed bacteria (16S rRNA gene sequencing), bile acid (LCMS) and metabolite (NMR) data were collected as described in the original publication[9]. Classification results of 16S data at the genus level are extracted for downstream analyses. Bacteria counts and bile acid intensities are transformed by log(x + 1); original NMR intensities are used. Target variables with more than 100 data points (non-missing value) are selected for downstream analysis. In total there are 49 target variables.

LonGP analysis

The following covariates have been selected for LonGP analysis. Continuous covariates include *age* (time from the start of experiment), *fmt* (time from the start of faecal microbiota transplant) and *abx* (time from the start of antibiotics). Discrete covariates include *batch* (faecal donor id) and *id* (chemostat model id). Interaction flags for *fmt* and *abx* are set to false. Default LonGP parameters are used. After LonGP deconvolution, the long-term trend is the sum of *age*, *batch* and *age*batch* components, while short-term trend is the *age*id* component. The parameter specification file is provided in Supplementary Data 1. The LonGP deconvolution visualizations are generated using gramm[10].

Weighted Gene Co-expression Network Analysis (WGCNA)

LonGP components in normalized scale are fed into the WGCNA R package[11]. The input was cleaned using the function “goodSamplesGenesMS” with default parameters. Then, soft-thresholding power for network construction was picked using the function “pickSoftThreshold”, where 8, 12, 12, are selected for long-term, short-term and raw data networks, respectively. Networks were constructed by function “blockwiseModules” with the following parameters: TOMType = "unsigned", minModuleSize = 2, reassignThreshold = 0, mergeCutHeight = 0.25, maxBlockSize = 5000. After that the function “TOMsimilarityFromExpr” and “exportNetworkToCytoscape” were used to export a network in edge and node list files in a format suitable for importing to Cytoscape[12]. Finally, edges with the top 150 absolute Pearson correlations were selected. Networks were visualized using R package igraph[13].

Batch cultures

Frozen glycerol stocks of *Ruminococcus gnavus* (DSMZ 108212), *Blautia obeum* and *Ruminococcus luti* (both isolated from human faecal material) were grown anaerobically at 37^{o}C (Don Whitley Scientific Limited, containing 80% N_2, 10% H_2 and 10% CO_2) for 24 hrs on brain-heart infusion (BHI; Sigma-Aldrich) agar supplemented with 3% L-cysteine hydrochloride monohydrate, D-(+)-Cellobiose, D-(+)-Maltose and 0.5% yeast extract. The identity of bacteria isolates was validated using Sanger sequencing (16S rRNA gene V1-V9), following bacterial DNA extraction (E.Z.N.A Omega Bio-Tek). Single bacteria colonies were inoculated in 5ml pre-reduced supplemented BHI, with added metabolites where appropriate (i.e. lithocholic acid at 1% for *Ruminococcus2* cultures) and incubated anaerobically for 24hrs at 37^{o}C. Samples were collected in triplicates at baseline and following 24hrs incubation and centrifuged at 12000*g* for 10 minutes at 4^{o}C. Supernatants were extracted and stored at -80^{o}C until further processing for bile acid profiling with liquid chromatography-mass spectrometry (LC-MS; Waters) and global metabolic profile using Nuclear Magnetic Resonance (NMR; Brucker 600MHz) using protocols as previously-described[14,9].

Nuclear magnetic resonance (NMR) data analysis

Supernatant from bacteria cultures, extracted as described above, were prepared for NMR analysis. Samples were diluted in NMR buffer (Supplementary Data 2) at 1:9 ratio. Samples were vortexed and centrifuged 20000g for 10mins at 4^{o}C for remove any possible precipitate. The supernatant was transferred in NMR tubes with an outer diameter of 5mm. Samples were processed using Bruker Avance IIII 600 MHz spectrometer (Bruker Corporation, Rheinstetten, Germany) with a standard NOESY pulse sequence, using protocols as previously-described[15]. Pre-processing of data was performed in MATLAB, and structual assignment was performed using data from literature, S-Base (Bruker), and in-house-databases[15].

LC-MS bile acid analysis

Samples were thawed thoroughly at room temperature for 1 hour and then diluted using a 1:1 water/ methanol mixture before 200 µL was added to a MS vial. LCA and 3-KCA standards were dissolved in methanol to obtain concentrations of 1.0 µM and 0.1 µM. Bile acid analysis and quantification was

performed using ACQUITY UPLC (Waters Corporation, Wilmslow, UK) coupled to a Xevo TQs mass spectrometer (Waters) equipped with an electrospray ionization source operating in negative ion mode (ESI-). LC-MS method was as described by Sarafian *et al*[16]. Peaks corresponding to 3-KCA and LCA were integrated using TargetLynx software (Waters) to obtain peak intensity values.

Results

Clostridioides difficile infection is a particularly interesting ecosystem to explore for this task, since disruption of the gut microbiome ecosystem by antibiotics is a major risk factor for the condition, and ecosystem restoration by FMT is a well-established treatment for the condition. In a previous longitudinal multi-omics study[9], faecal samples from human donors were implanted in artificial gut models (Fig. 1) to simulate the gut environment *in vitro* to study the effect of FMT on inhibiting the growth of *C. difficile*. Samples were collected from the models on alternate days, and bacterial abundance (16S rRNA gene sequencing), bile acid and metabolite profiles were generated from these samples, which were called target variables in the analysis.

We first used LonGP[7] to deconvolute each target variable into various covariate associated components, such as the global trend shared by all models, the batch specific trend shared by models that use faecal samples from the same donor, the FMT trend shared by cases and short-term trend that is model-specific. This analysis has successfully uncovered that valerate concentration first decreases after the introduction of antibiotics and only increases after FMT (Fig. 2a), which confirms the serendipitious main finding of the original publication and helps to validate our approach.

Based on the LonGP regression results, we extracted the long-term and short-term trends to construct separate bacteria-metabolite interaction networks using WGCNA[11]. The long-term interaction network (Fig. 2b, Supplementary Data 3) led us to focus on *Blautia*, which was not involved in any interaction in the raw data interaction network (Supplementary Data 3). These negative interactions (*Blautia* with ethanol, 5-aminovalerate and succinate) were presumed to be weak and hard to detect using direct associations. Batch culture experiments incubating *Blautia obeum* in brain-heart infusion (BHI) medium validated the interaction with succinate (Fig. 2c), but not interactions with ethanol and 5-aminovalerate (no change in all replicates). This interaction between *Blautia* and succinate would not have been expected without the network analysis, and to our knowledge has not been previously-described.

The short-term interaction network (Fig. 2d, Supplementary Data 3) identified new transient interactions compared with the raw data network (Fig. 2e), e.g. *Ruminococcus2* positively correlate with 3-ketocholanic acid (KCA), whose LonGP deconvolutions were provided in Supplementary Fig. 1–2. In batch culture experiments (Fig. 2f) incubating two representative human microbiome members of *Ruminococcus2* (*R. gnavus* and *R. luti*) with lithocholic acid (LCA), we found that KCA was generated by both *R. gnavus* and *R. luti*. In particular, KCA was detected in the batch culture of *R. gnavus* in media with LCA at baseline, which is surprisingly fast. This was not contamination since the LCA control (identical medium without *R. gnavus*) contained no KCA, which suggested KCA must be generated during the set up

of the batch culture at 0 hour (took ~ 10 minutes). Reduced KCA was detected in the *R. gnavus +* LCA batch culture at 24 hours compared to 10 minutes, suggesting further ongoing bacterial metabolism. No KCA was detected in the baseline *R. luti/* LCA batch culture but was detectable by 24 hrs; this suggests variability in the dynamics between *Ruminococcus2* members at which LCA is metabolised to KCA.

Other findings from the networks supported established knowledge about bacteria-metabolite interactions; for example, a positive correlation was observed between the genus *Clostridium* XIVa and lithocholic acid in the short-term network (Supplementary Fig. 3a), with this being one of the only genera containing bacteria able to bioconvert primary to secondary bile acids[17]. The placement of trimethylamine (TMA) in short- and long-term networks (Supplementary Fig. 3a,b) is supported by the prior recognition that TMA is one of the identified metabolites most strongly restored by FMT for *C. difficile* infection[1] (CDI). Similarly, the positive interaction between glycodeoxycholic acid (GDCA) and *Clostridium* XI (the genus containing *C. difficile*) seen in short and long-term networks (Supplementary Fig. 3c,d) supports the prior recognition of GDCA in promoting germination of spore forming bacteria including *C. difficile*[18]. A further notable network feature is the loss of *Faecalibacterium* after antibiotics and restoration after FMT (Supplementary Fig. 4), especially of note given the previous observation of the absence of *F. prausnitzii* in inflammatory intestinal conditions including CDI and inflammatory bowel disease[19], and the anti-inflammatory properties associated with this bacterium.

Discussion

As a proof-of-concept, we demonstrate there is a high potential for large scale high-throughput microbiome studies to screen for interesting host/microbe/metabolite interactions. Two out of four candidate interactions have been validated (accuracy is 50%), which reflects the reliability of the networks. The identified transient interaction between *R. gnavus* and KCA suggests that interactions could be extremely fast, which hints the short-term interaction network is a valuable source to look for fast interactions. In our analysis, we define the short- and long-term effects to be fast and slow changing in all chemostat models. Changes related with antibiotic treatment and FMT that are specific to the case group are excluded in the long-term trend. These effects, however, contain important information about bacteria/metabolite dynamics and worth further investigation, e.g. valerate (Fig. 2a) and *Faecalibacterium* (Supplementary Fig. 4).

Conclusions

In summary – via validation of established interactions and through culture-based confirmations of putative interactions - we here provide a new analytical method to identify transient and stable bacteria-metabolite interactions from longitudinal multi-omics study with complex experimental design. We feel that deconvoluting the raw data into independent sources makes the inferred networks better reflect the underlying interaction mechanisms and dynamics of the microbiota. We anticipate this analytical

approach to be useful for a wide range of microbiome studies. In particular, identified short-term interactions might greatly benefit and accelerate biosynthesis and culturomics applications[20].

Abbreviations

BHI
brain-heart infusion
CDI
Clostriodioides difficile infection
FMT
faecal microbiota transplant
GDCA
glycodeoxycholic acid
KCA
3-ketocholanic acid
LCA
lithocholic acid
LCMS
Liquid chromatography–mass spectrometry
NMR
Nuclear magnetic resonance
TMA
trimethylamine
WGCNA
Weighted Gene Co-expression Network Analysis

Declarations

Ethics approval and consent to participate (Not applicable)

Consent for publication (Not applicable)

Availability of data and material

All data are included in the supplementary data.

Competing interests

None declared.

Funding

This work is supported by the National Key Research and Development Program of China, Stem Cell and Translational Research (2017YFA0106800) and the National Science Fund for Excellent Young Scholars (81722004) for L.C., National Natural Science Foundation of China (31870107, 32070122) to J. W., and Academy of Finland (335858) to L. Cheng.

The Department of Metabolism, Digestion and Reproduction at Imperial College London receives funding from the National Institute of Health Research (NIHR) Biomedical Research Centre (BRC) based at Imperial College London and Imperial College Healthcare NHS Trust. BHM was the recipient of a Medical Research Council (MRC) Clinical Research Training Fellowship (MR/R000875/1) and is the recipient of an NIHR Academic Clinical Lectureship (CL-2019-21-002). JAKM is the recipient of a Wellcome Trust Institutional Strategic Support Fund (ISSF) fellowship.

Authors' contributions

L. Cheng, J. W. and B. H. M. conceptualized the project. L. Cheng oversaw the whole project. D. Z. and L. C. performed the network analysis. S. G. and G. B. performed the NMR and LCMS data analysis. C. D. cultured the bacteria. J. A.K. M. and J.R. M. provide insights in interpreting the results.

Acknowledgements

L. Cheng acknowledges the computational resources provided by the Aalto Science-IT project.

Authors' information (optional)

Dan Zhang, Benjamin H Mullish and Jinfeng Wang are co-first authors, Julian R. Marchesi and Lu Cheng are co-corresponding authors.

References

1. Jansma, J. & Aidy, S. El. Understanding the host-microbe interactions using metabolic modeling. Microbiome **9**, 16 (2021).
2. Sung, J. *et al.* Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. Nat. Commun. **8**, 15393 (2017).
3. Mars, R. A. T. *et al.* Longitudinal Multi-omics Reveals Subset-Specific Mechanisms Underlying Irritable Bowel Syndrome. Cell **182**, 1460–1473.e17 (2020).
4. Metwaly, A. *et al.* Integrated microbiota and metabolite profiles link Crohn's disease to sulfur metabolism. Nat. Commun. **11**, 4322 (2020).
5. Leonard, M. M. *et al.* Microbiome signatures of progression toward celiac disease onset in at-risk children in a longitudinal prospective cohort study. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 29 (2021).
6. Feng, Q. *et al.* Time series analysis of microbiome and metabolome at multiple body sites in steady long-term isolation confinement. Gut **70**, 1409–1412 (2021).

7. Cheng, L. *et al.* An additive Gaussian process regression model for interpretable non-parametric analysis of longitudinal data. Nat. Commun. **10**, 1–11 (2019).
8. Martinez-Gili, L. *et al.* Understanding the mechanisms of efficacy of fecal microbiota transplant in treating recurrent Clostridioides difficile infection and beyond: the contribution of gut microbial-derived metabolites. Gut Microbes **12**, 1810531 (2020).
9. McDonald, J. A. K. *et al.* Inhibiting Growth of Clostridioides difficile by Restoring Valerate, Produced by the Intestinal Microbiota. Gastroenterology **155**, 1495–1507.e15 (2018).
10. Morel, P. Gramm: grammar of graphics plotting in Matlab. J. Open Source Softw. **3**, 568 (2018).
11. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics **9**, 559 (2008).
12. Paul Shannon, 1 *et al.* Cytoscape: A Software Environment for Integrated Models. *Genome Res.* **13**, 2498–2504 (2003).
13. Csardi, G. & Nepusz, T. The igraph software package for complex network research. InterJournal Complex Syst. 1695 (2006).
14. Mullish, B. H. *et al.* Microbial bile salt hydrolases mediate the efficacy of faecal microbiota transplant in the treatment of recurrent Clostridioides difficile infection. Gut **68**, 1791–1800 (2019).
15. Dona, A. C. *et al.* A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. Comput. Struct. Biotechnol. J. **14**, 135–153 (2016).
16. Sarafian, M. H. *et al.* Bile Acid Profiling and Quantification in Biofluids Using Ultra-Performance Liquid Chromatography Tandem Mass Spectrometry. Anal. Chem. **87**, 9662–9670 (2015).
17. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature **517**, 205–208 (2015).
18. Tanaka, M., Onizuka, S., Mishima, R. & Nakayama, J. Cultural isolation of spore-forming bacteria in human feces using bile acids. Sci. Rep. **10**, (2020).
19. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature **569**, 655–662 (2019).
20. Weiss, A. S. *et al.* In vitro interaction network of a synthetic gut bacterial community. ISME J. (2021).

Figures

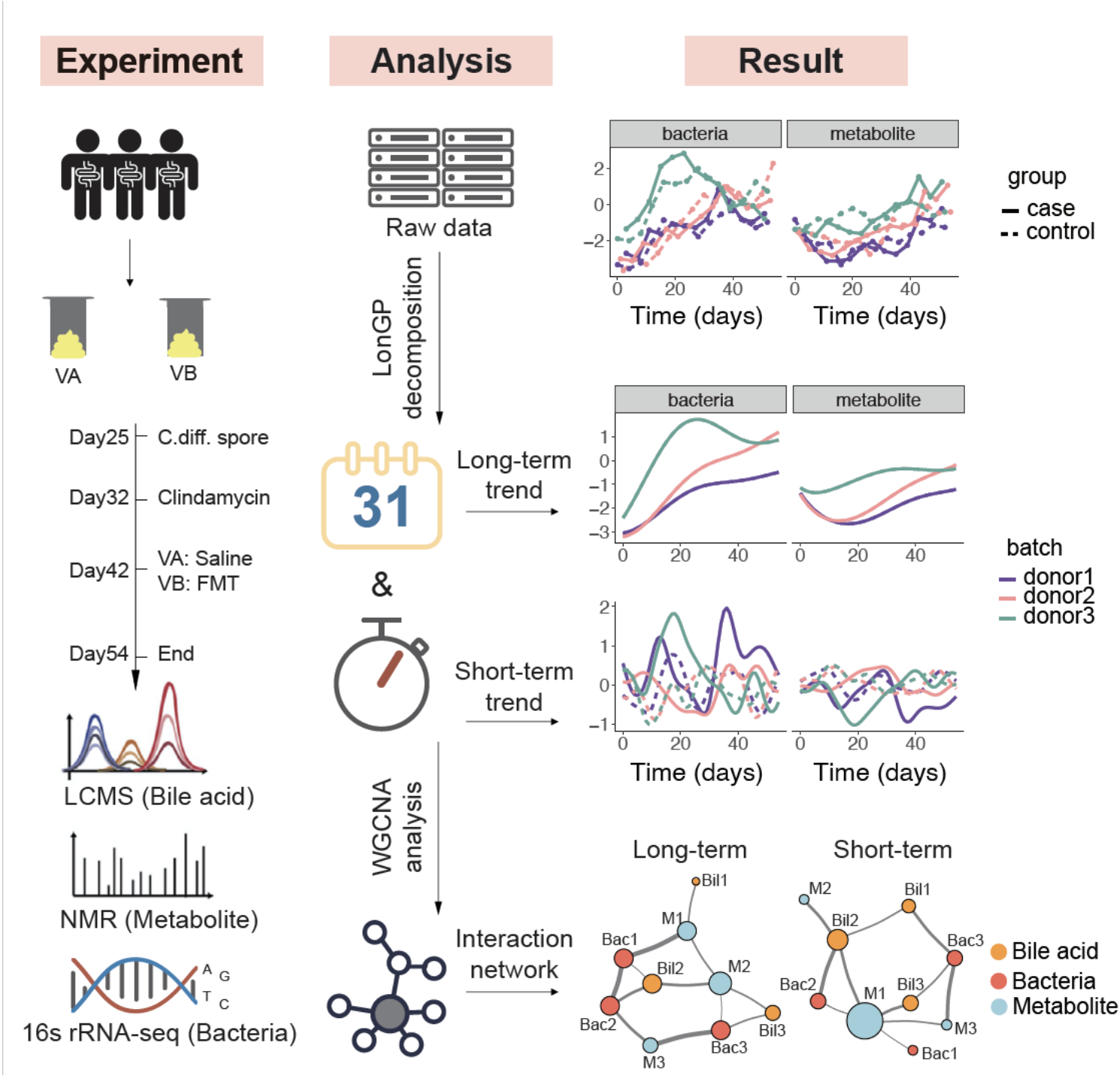

Figure 1

Analysis workflow.

(**Experiment)** Faecal samples are collected from 3 human donors. Each sample is implanted in two chemostat models, where *Clostridioides difficile* spores, antibiotics (clindamycin), faecal microbiota transplant (FMT, case) or Saline (control) were introduced at day 25, 32 and 42. Samples were collected from chemostat models every 2 days and measurements of bile acids, metabolites and microbiome were taken for these samples.

(**Analysis**) Raw microbiome/metabolite data was first deconvoluted into long-term & short-term trends using LonGP, which were then fed into WGCNA (Weighted correlation network analysis) to generate separate interaction networks.

(**Result**) It can be seen that the long-term and short-term trends exhibit positive and negative associations, respectively. Note that some deconvoluted components such as case specific trends are not included in the long-term trend.

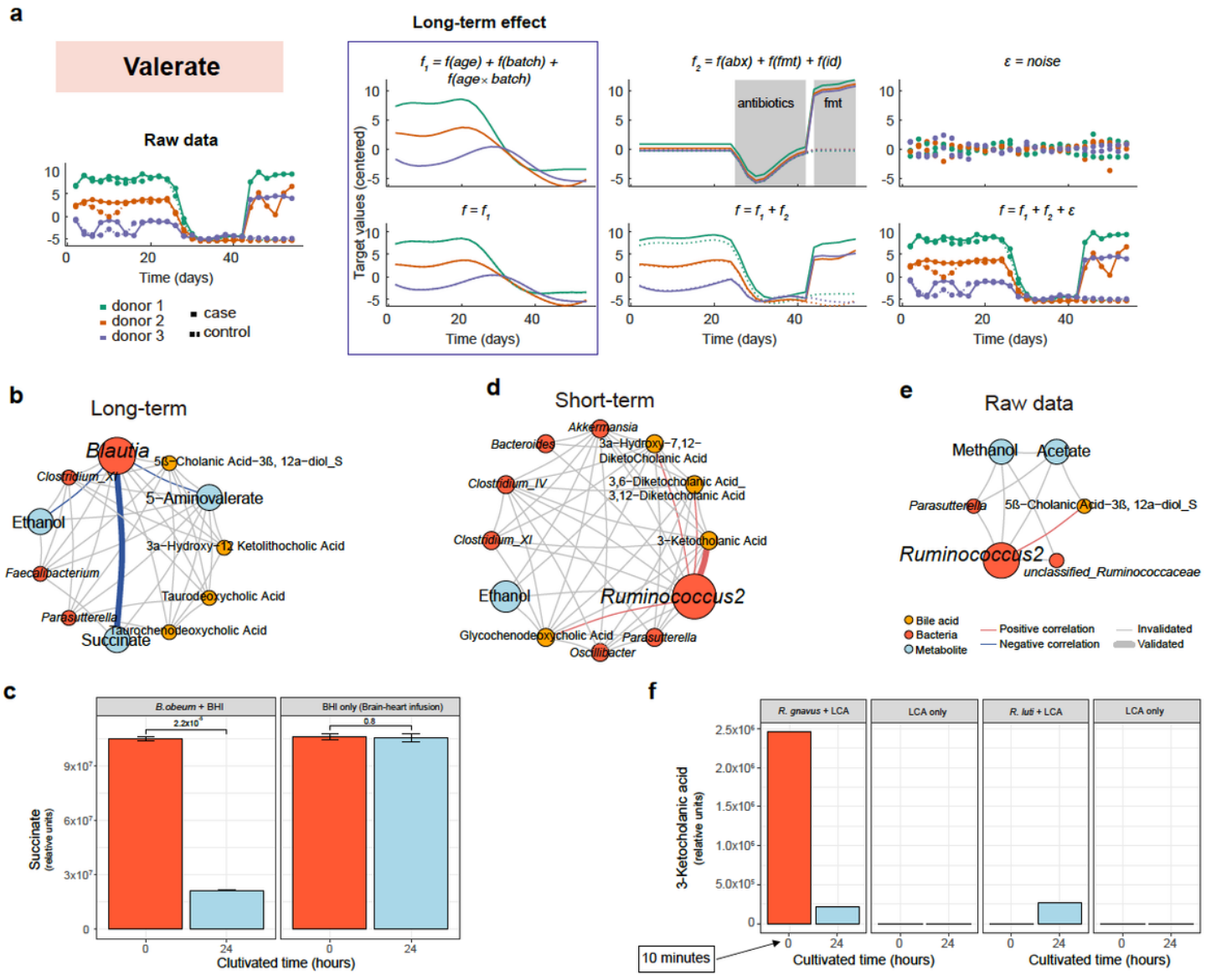

Figure 2

LonGP deconvolution and bacteria-metabolite interaction networks based on raw data, long-term and short-term trends. The strongest 150 interactions (ranked by absolute Pearson correlation) are retained for each network. Only subnetworks of the chosen bacteria and their neighbouring nodes are shown for

visualization purposes. Positive and negative correlations are indicated by red and blue edges, respectively.

(a) LonGP deconvolution of NMR data of valerate. The top row are different components, while the bottom row are the cumulative sums of these components. Shaded areas indicate antibiotics and FMT treatment periods. Note that no short-term effect is detected for valerate by LonGP, i.e. short-term effect is not significant.

(b) Long-term network centered on *Blautia*, which suggests interactions with ethanol, 5-aminovalerate and succinate. Note that raw data network centered on *Blautia* cannot be displayed since it is not involved in any top 150 interaction.

(c) Validation results of culturing *Blautia obeum* in brain-heart infusion (BHI) medium for 24 hours. The Y-axis is quantified intensities of NMR peaks. Each bar with error bars represents measurements of all available technical replicates, where t-test is performed between 0 and 24 hours. Significant decreases of succinate in 24 hours can be observed by comparing cases (*B. obeum* + BHI) versus the control (BHI only).

(d) Short-term network centered on *Ruminococcus2*. Compared with the raw data network, this network identifies novel interactions with 3-ketocholanic acid, glycochenodeoxycholic acid, 3,6-diketocholanic acid/ 3,12-diketocholanic acid, and 3a-hydroxy-7,12-diketocholanic acid.

(e) Raw data networks centered on *Ruminococcus2*.

(f) Validation results of culturing *R. luti* and *R. gnavus* with lithocholic acid (LCA) for 24 hours. Y-axis is the quantified intensity. Note that there is only one replicate per condition. It is clear that *R. luti* has generated 3-ketocholanic acid (KCA). By comparing the case with the control at 0 hour, it can be seen that *R. gnavus* has generated KCA. Note that it takes around 5 minutes to prepare the sample at 0 hour, so the result suggests *R. gnavus* can generate KCA at an ultra-high speed, which is probably converted into some other metabolites at 24 hour.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementData1LonGPresults.xlsx
- SupplementData2NMRandLCMSvalidationresults.xlsx
- SupplementData3WGCNAtop150networksrawlongtermshortterm.cys
- SupplementData4LCMSNMRrawdata.zip
- 1.png

- 2.png
- 3.png
- 4.png