

Hidden link in gut–joint axis: gut microbes promote rheumatoid arthritis at early stage by enhancing ascorbate degradation

With great interest, we read the review article by Agus *et al*, which suggested that gut microbiome alterations could affect metabolic homeostasis.¹ Moreover, gut microbiome alterations in concert with metabolites perturbation could contribute to the early development of rheumatoid arthritis (RA).² We thus conducted a three-pronged association study³ on multiomics datasets to detect the potential microbiome–metabolites–arthritis link.

We integrated multiomics datasets including gut metagenomics, clinical phenotypes and metabolites of blood and knee-joint synovial fluid from 122 participants in the healthy group (n=27), osteoarthritis (OA) group

(n=19) and RA group (n=76), using a three-pronged association framework (figure 1, online supplemental material).³ Metagenomic genes were collapsed into metagenomic species (MGS)^{3,4} and grouped into KEGG functional modules (figure 1A).³ Additionally, the co-abundant metabolites were categorised into metabolite clusters using WGCNA framework (figure 1A).³ The functional modules associated with clinical phenotypes (eg, types of arthritis and levels of cytokines) were further identified and the cross-domain associations between these modules and metabolite clusters were assessed (figure 1B).³ Furthermore, the leave-one-out analysis was performed to determine the MGS that particularly contributed to the observed linkage between functional modules and clinical phenotypes (figure 1C).³

We found that gut microbial functionality in ascorbate degradation (KEGG module: M00550) was positively correlated with the types of arthritis (healthy=0, OA=1, RA=2, $p_{\text{Wilcox}}=2.15 \times 10^{-4}$) and the levels of proinflammatory cytokines TNF- α (tumour necrosis factor- α , $p_{\text{Wilcox}}=6.59 \times 10^{-4}$) and IL-6 (interleukin-6, $p_{\text{Wilcox}}=1.12 \times 10^{-3}$). Ascorbate (vitamin C) was previously reported to prevent the development of inflammatory arthritis,⁵ possibly through facilitating collagen synthesis, moderating autoimmune responses and ameliorating inflammation.⁶ Additionally, the patients with RA are usually ascorbate deficient and require high-dose supplementation to maintain an acceptable plasma level of ascorbate.⁷ In this study, the functional module of ascorbate degradation was observed to positively correlate with the blood metabolite cluster MB02 ($p_{\text{Wilcox}}=6.90 \times 10^{-3}$), which was represented by the level of palmitic acid (kME (eigengene-based connectivity) =0.911, kIN (intramodular connectivity) =3.46, online supplemental table 1) that acts as a proinflammatory factor, upregulating IL-6 secretion by human chondrocytes and fibroblast-like synovial cells in inflammatory arthritis.⁸ Furthermore, we found that *Escherichia coli* and *Streptococcus bovis* were the driving species for the observed linkage between ascorbate

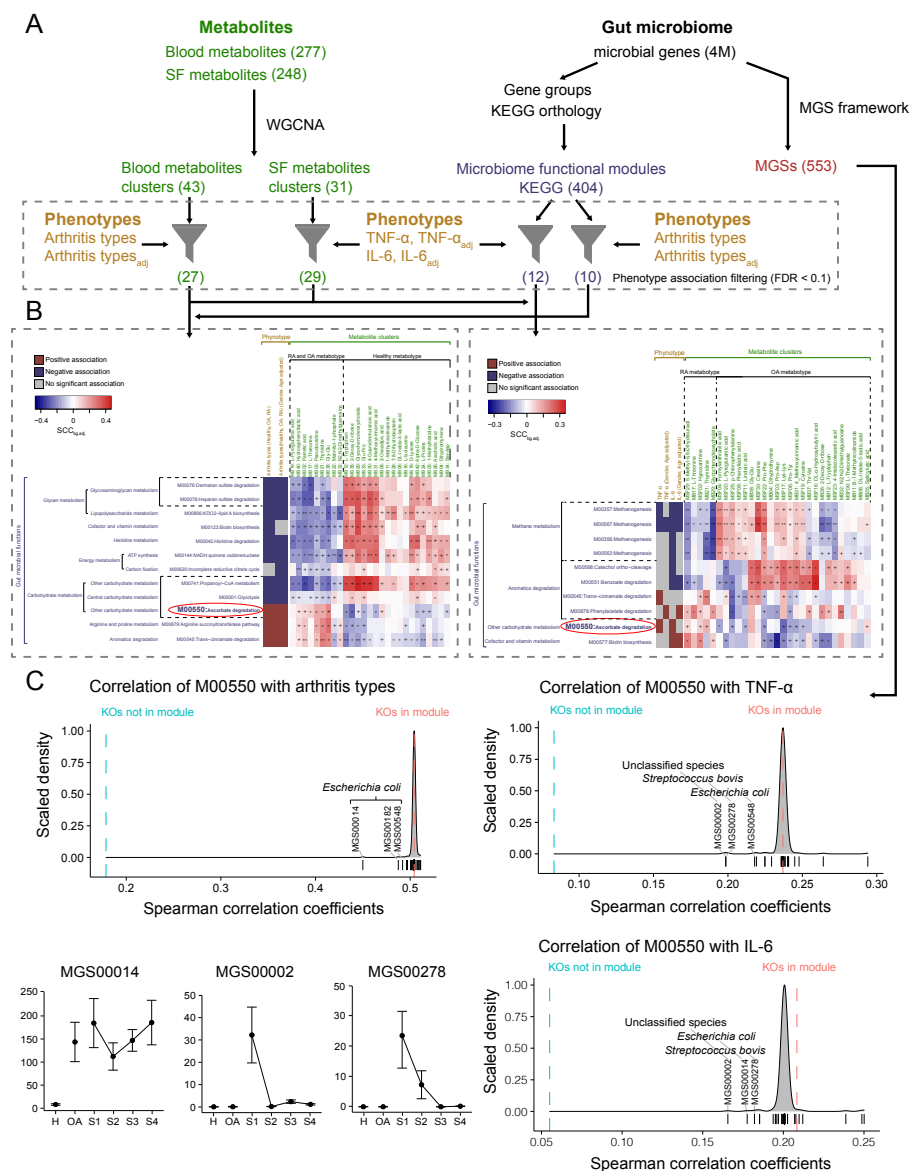


Figure 1 Overview of the three-pronged association framework integrating multiomics datasets.

(A) Metabolites are summarised as co-abundance clusters, and microbial genes are grouped into KEGG modules and MGS, which are further filtered for statistically positive or negative associations (based on Spearman correlation) with the clinical phenotypes. The association analyses were divided by using healthy, OA and RA samples for arthritis types and using OA and RA samples for cytokine levels. The number in brackets represent the number of metabolites/metabolite clusters/microbial genes/KEGG modules/MGS in each analytical module. (B) The filtered features are further used for cross-domain association analyses. For each analysis, the left panel shows the significant associations (Mann-Whitney U test FDR<0.1) between KEGG modules and clinical phenotypes, and colour indicates significantly positive association (red), significantly negative association (blue) or insignificant association (grey). The right panel shows the associations between KEGG modules and metabolite clusters, and the colour represents the median Spearman correlation coefficient (SCC) of metabolite clusters with KEGG orthologies (KOs) in KEGG module minus those with KOs not in KEGG module. Mann-Whitney U test FDRs are denoted: *FDR<0.1; **FDR<0.01; ***FDR<0.001. (C) The MGS that particularly contributed to the observed linkage between functional modules and clinical phenotypes. Three density plots: Dashed line represents the median SCC of the phenotypes with KOs in M00550 (red) and all other KOs (blue). Density plot shows the median SCC of the phenotypes with KOs in M00550, when a given MGS (indicated by short vertical lines) has been excluded from the analysis. The bottom-left dot plots show the mean \pm SEM of the top three driving MGS abundances among patients at each stage of disease development, with the four RA stages connected to display the variance. FDR, false discovery rate; IL-6, interleukin-6; MGS, metagenomic species; OA, osteoarthritis; RA, rheumatoid arthritis; TNF- α , tumour necrosis factor- α .

degradation⁹ and the arthritis types or the cytokines levels of TNF- α and IL-6 (figure 1C). Subsequently, we grouped patients with RA by four stages according to the comprehensive scores in rheumatoid diagnostic criteria,¹⁰ as RASI: 6–7, RASII: 8, RASIII: 9 and RASIV: 10 (online supplemental table 2). We observed that both *E. coli* and *S. bovis* were prevalent at RA stage I (RASI), while *S. bovis* was depleted after RASI or in the OA group. It suggested *S. bovis* mainly functioned at the early stage of RA, while *E. coli* might be crucial throughout the entire developmental stages of RA and OA. Taken together, we speculate that *E. coli* and *S. bovis* could facilitate ascorbate degradation and thus promote proinflammatory responses that facilitate the development of inflammatory arthritis.

Overall, we demonstrate that gut microbiota could promote RA progression via enhancing ascorbate degradation and provide a potential approach to prevent the development of arthritis through interfering gut–joint axis. The results of this study could be prospected in following contexts: First, our study provides a reservoir of the potential microbiome–metabolites–arthritis links as a reference of gut–joint axis for future studies. Second, the findings supplement the potential mechanisms related to metabolic perturbation through which gut microbiome promotes arthritis.¹² Third, considering the inflammatory pathways of arthritis were revisited in COVID-19,¹¹ it deserves further investigations whether microbiome–ascorbate–inflammation link of this study could contribute to the treatment of COVID-19.

Yan Zhao¹,¹ Mingyue Cheng²,
Liang Zou¹,¹ Luxu Yin¹,¹ Chaofang Zhong,²
Yuguo Zha²,² Xue Zhu²,² Lei Zhang³,³
Kang Ning²,² Jinxiang Han¹

¹First Affiliated Hospital of Shandong First Medical University, Biomedical Sciences College & Shandong Medical Biotechnology Centre, Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan, China

²Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Key Laboratory of Bioinformatics and Molecular-imaging, Center for Artificial Intelligence Biology, Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China

³Microbiome-X, National Institute of Health Data Science of China & Institute for Medical Dataology, Department of Biostatistics, School of Public Health, CheeLo College of Medicine, Shandong University, Jinan, China

Correspondence to Professor Lei Zhang, Shandong University, Jinan, China; zhanglei7@sdu.edu.cn, Professor Kang Ning, Huazhong University of Science and Technology, Wuhan, China; ningkang@hust.edu.cn and Professor Jinxiang Han,

Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan, China; jxhan@sdfmu.edu.cn

Correction notice This article has been corrected since it published Online First. The figure has been replaced for clarity.

Contributors YZ and MC designed the study, conducted the data analysis and wrote the manuscript. YZ, MC, LZ, LY, CZ, YZ and XZ collected the samples, conducted the experiments and participated in data analysis. LZ, KN and JH supervised the study and revised the manuscript.

Funding This study was funded by National Natural Science Foundation of China (grant numbers 32071465, 82003766, 31871334, and 31671374), Academic Promotion Project of Shandong First Medical University (grant number 2019LJ001) and Ministry of Science and Technology's National Key Research and Development Programme (grant number 2018YFC0910502).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the Ethics Committee of the First Affiliated Hospital of Shandong First Medical University (No. 2017-02).

Provenance and peer review Not commissioned; externally peer reviewed.

Supplementary material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.



OPEN ACCESS

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2021-325209>).

YZ and MC contributed equally.



To cite Zhao Y, Cheng M, Zou L, *et al.* Gut 2022;71:1041–1043.

Received 19 May 2021

Accepted 3 July 2021

Published Online First 8 July 2021

Gut 2022;71:1041–1043. doi:10.1136/gutjnl-2021-325209

ORCID iDs

Yan Zhao <http://orcid.org/0000-0002-2816-8951>
Mingyue Cheng <http://orcid.org/0000-0003-1243-5039>
Liang Zou <http://orcid.org/0000-0003-0899-2318>
Luxu Yin <http://orcid.org/0000-0003-2374-9965>
Yuguo Zha <http://orcid.org/0000-0003-3702-9416>
Xue Zhu <http://orcid.org/0000-0002-7337-6186>
Lei Zhang <http://orcid.org/0000-0003-4284-2604>
Kang Ning <http://orcid.org/0000-0003-3325-5387>
Jinxian Han <http://orcid.org/0000-0001-9269-4553>

REFERENCES

- 1 Agus A, Clément K, Sokol H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut* 2021;70:1174–82.
- 2 Zaiss MM, Joyce Wu H-J, Mauro D, *et al.* The gut-joint axis in rheumatoid arthritis. *Nat Rev Rheumatol* 2021;17:224–37.
- 3 Pedersen HK, Forslund SK, Gudmundsdottir V, *et al.* A computational framework to integrate high-throughput ‘-omics’ datasets for the identification of potential mechanistic links. *Nat Protoc* 2018;13:2781–800.
- 4 Nielsen HB, Almeida M, Juncker AS, *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* 2014;32:822–8.
- 5 Pattison DJ, Silman AJ, Goodson NJ, *et al.* Vitamin C and the risk of developing inflammatory polyarthritis: prospective nested case-control study. *Ann Rheum Dis* 2004;63:843–7.
- 6 Carr AC, Maggini S. Vitamin C and immune function. *Nutrients* 2017;9:1211.
- 7 Abrams E, Sandson J. Effect of ascorbic acid on rheumatoid synovial fluid. *Ann Rheum Dis* 1964;23:295–9.
- 8 Frommer KW, Schäffler A, Rehart S, *et al.* Free fatty acids: potential proinflammatory mediators in rheumatic diseases. *Ann Rheum Dis* 2015;74:303–10.
- 9 Yew WS, Gerlt JA. Utilization of L-ascorbate by *Escherichia coli* K-12: assignments of functions to products of the *yji*-sga and *yia*-sgb operons. *J Bacteriol* 2002;184:302–6.
- 10 Aletaha D, Neogi T, Silman AJ. Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League against rheumatism collaborative initiative. *Arthritis Rheum* 2010;2010:2569–81.
- 11 Schett G, Manger B, Simon D, *et al.* COVID-19 revisiting inflammatory pathways of arthritis. *Nat Rev Rheumatol* 2020;16:465–70.

Supplementary information

A hidden link in gut-joint axis: Gut microbes promote rheumatoid arthritis at early stage by enhancing ascorbate degradation

Yan Zhao,^{1#} Mingyue Cheng,^{2#} Liang Zou,¹ Luxu Yin,¹ Chaofang Zhong,² Yugo Zha,² Xue Zhu,² Lei Zhang,^{3,4*} Kang Ning,^{2*} Jinxiang Han^{1*}

¹Shandong Medicine and Health Key Laboratory of Rheumatism, Shandong Key Laboratory of Rheumatic Disease and Translational Medicine, Department of Rheumatology and Autoimmunology, Shandong Provincial Qianfoshan Hospital, First Affiliated Hospital of Shandong First Medical University. Jinan 250014, Shandong, China

²Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Key Laboratory of Bioinformatics and Molecular-imaging, Center of AI Biology, Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

³Microbiome-X, National Institute of Health Data Science of China & Institute for Medical Dataology, Cheeloo College of Medicine, Shandong University, Jinan 250014, Shandong, China

⁴Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250014, Shandong, China

***Correspondence to** Dr Jinxiang Han, E-mail: jxhan@sdfmu.edu.cn; Dr Kang Ning, E-mail: ningkang@hust.edu.cn; Dr Lei Zhang, E-mail: zhanglei7@sdu.edu.cn.

[#]These authors contributed equally to this work

29 Sample description

30 A total of 122 fecal and 122 serum samples were collected from 122 outpatients from
31 the Shandong Provincial Qianfoshan Hospital (Jinan, Shandong, China). These
32 outpatients included 27 healthy individuals, 19 patients with osteoarthritis (OA), and
33 76 patients with rheumatoid arthritis (RA). Subsequently, the fecal samples were
34 sequenced and the serum samples were used to examine serum metabolites and
35 inflammatory cytokines. Serum inflammatory cytokines TNF- α and IL-6 were
36 quantified by the MESO SCALE DISCOVERY (MSD®) Quick Plex S600MM
37 multiplex assay. The cytokine levels of healthy individuals were extremely low and not
38 available. In addition, 95 knee-joint synovial fluid samples were collected from the RA
39 and OA patients to examine synovial fluid metabolites. Both serum and synovial fluid
40 metabolites were examined by UHPLC-MS/MS.

41 All of the participants were at fasting status during sample collection in the morning.

42 The participants were recruited in this study following the standards shown below:

- 43 1. Healthy individuals in good health condition with no gastrointestinal diseases, such
44 as diarrhea, constipation, and hematochezia, in the recent one month, no
45 hepatobiliary system diseases, no history of gastrointestinal tumors or inflammatory
46 diseases, no serious heart, liver, kidney, lung, brain or other organ disorders, no
47 infections, chronic diseases, or antibiotic treatment;
- 48 2. Healthy individuals had not taken any acid inhibitors, gastrointestinal motility drugs,
49 antibiotics, or living bacteria products such as yogurt in the recent one month;
- 50 3. Healthy individuals with no history or family history of mental illness, and no
51 history of gastrointestinal surgery;
- 52 4. RA/OA individuals with no other co-morbidity.

53 Metagenome sequencing and data processing

54 Whole-genome shot-gun sequencing of fecal samples were carried out on the Illumina
55 HiSeq X Ten. All samples were paired-end sequenced with a 150-bp read length. After

quality control, the paired-end reads were assembled into contigs using MEGAHIT (version 1.2.6)¹ with the minimum contig length set at 500 bp. The open reading frames (ORFs) were predicted from the assembled contigs using Prodigal (version 2.6.3)² with default parameters. The ORFs of <100 bp were removed. The ORFs were then clustered to remove redundancy using Cd-hit (version 4.6.6)³ with a sequence identity threshold set at 0.95 and the alignment coverage set at 0.9, which resulted in a catalog of 4,047,645 non-redundant genes. The non-redundant genes were then collapsed into metagenomic species (MGS)^{4,5} and grouped into KEGG functional modules.⁴

Identification of MGS

High-quality reads were mapped to the catalog of non-redundant genes using Bowtie 2 (version 2.2.9)⁶ with default parameters. The abundance profile for each catalogue gene was calculated as the sum of uniquely mapped sequence reads, using 19M sequence reads per sample (downsized). The co-abundance clustering of the 4,047,645 genes was performed using canopy algorithm (<http://git.dworzynski.eu/mgs-canopy-algorithm>),⁵ and 553 gene clusters that met the previously described criteria⁵ and contained more than 700 genes were referred to as MGS. MGS present in at least 4 samples were used for the following analysis. The abundance profiles of MGS were determined as the medium gene abundance throughout the samples. MGS were taxonomically annotated as described by Nielsen *et al.*⁵ and each MGS gene was annotated by sequence similarity to NCBI bacterial genome (BLASTN, E-value < 0.001)

Annotation of KEGG modules

The catalog of the non-redundant genes was functionally annotated to KEGG database (release 94.0) by KofamKOALA (version 1.3.0).^{7,8} The produced KEGG Orthologies (KOs) were mapped to the KEGG modules annotation downloaded on August 1, 2020 from the KEGG BRITE database. KOs present in at least 4 samples were used for the following analysis. The KO abundance profile was calculated by summing the

82 abundances of genes that were annotated to each KO.

83 **Clustering of co-abundant metabolites**

84 Co-abundant metabolites in serum or synovial fluid were identified using the R package
85 WGCNA⁹. As recommended by Pedersen *et al.*,⁴ a signed network and biweighted mid-
86 correlation were used for clustering with the soft threshold $\beta = 8$ for both serum and
87 synovial fluid metabolites. The minimum cluster size was set as 3. Similar clusters were
88 subsequently merged if the biweight mid-correlation between the cluster's eigen
89 vectors exceeded 0.8 for both serum and synovial fluid metabolites. The kIN of a
90 metabolite was calculated by summing connectivity with all other metabolites in the
91 given metabolite cluster. The kME was determined by the bicor-correlation between
92 the metabolite profile and module eigenvector. Both kIN and kME were used to
93 measure the intramodular hub-metabolite status.

94 **Cross-domain association analyses**

95 The clinical phenotypes, including types of arthritis (Healthy = 0, OA = 1, RA = 2) and
96 the levels of pro-inflammatory cytokines TNF- α and IL-6, were used in the association
97 analysis. TNF- α and IL-6 were selected based on their potentials to act as the
98 therapeutic targets for RA treatment.^{10 11} The associations between clinical phenotypes
99 and KEGG modules/metabolites clusters were determined through evaluating if the
100 Spearman correlations of the phenotype with the abundances of KOs/metabolites in the
101 given KEGG module/metabolite clusters were significantly higher or lower (Mann-
102 Whitney U-test FDR < 0.1) than with the abundances of all other KOs/metabolites. The
103 phenotypes adjusted by age and gender were also tested. Moreover, the union set of the
104 significant associations between KEGG modules and phenotypes/phenotypes adjusted
105 by age and gender, and the intersect set of the significant associations between
106 metabolites clusters and phenotypes/phenotypes adjusted by age and gender, were used
107 for the following association analysis. The associations between metabolite clusters and

108 KEGG modules were determined through evaluating if the Spearman correlations of
109 the eigen vectors of the metabolite clusters with the abundances of KOs in the given
110 KEGG module were significantly higher or lower (Mann–Whitney U-test FDR < 0.1)
111 than with the abundances of all other KOs/metabolites.

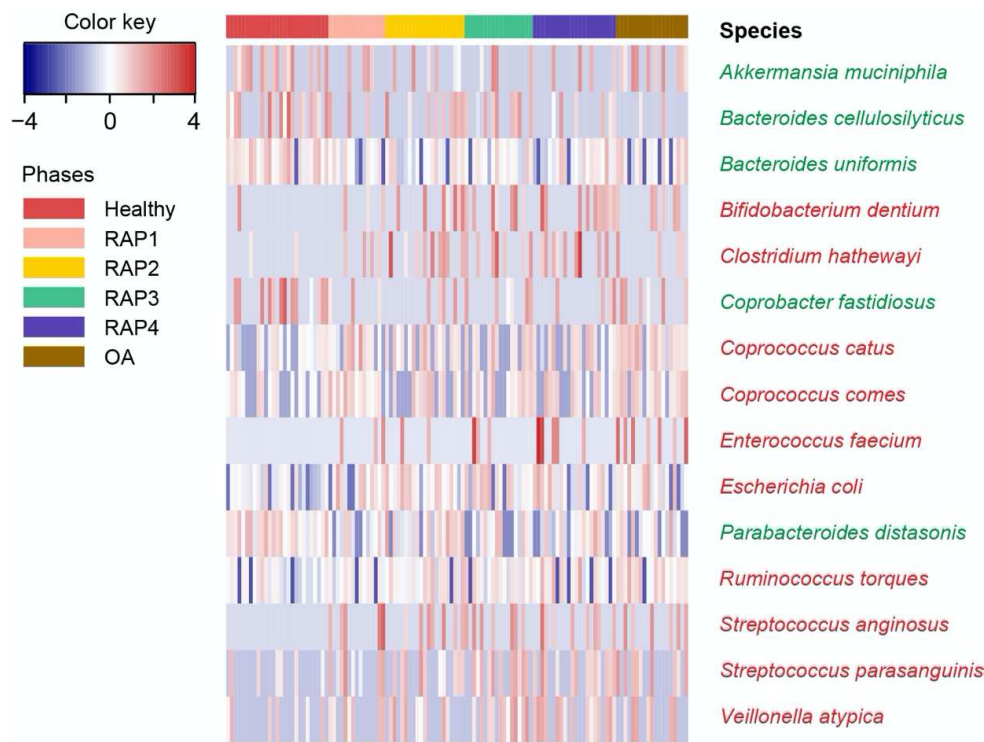
112 **Leave-one-out analysis**

113 Leave-one-out analysis was used to identify the specific MGS driving the observed
114 associations between KEGG module M00550 and the clinical phenotypes, including
115 the types of arthritis or the levels of pro-inflammatory cytokines TNF- α and IL-6. The
116 calculation of the KO abundance was iterated excluding the genes from a different MGS,
117 in each iteration. The effect of a given MGS on a specified association was defined as
118 the change in median Spearman correlation coefficient between KOs and clinical
119 phenotypes when genes from the respective MGS were left out, as previously
120 described.^{4 12}

121 **Taxonomic identity of differentially present microbes across conditions**

122 MetaPhlAn2¹³ was used to generate species profiles. Species that were present in less
123 than 10% samples were excluded. Supplementary Figure 1 displays the union set of the
124 species (n=15) with significantly different abundances (Mann–Whitney U-test FDR <
125 0.05) between the healthy and RA groups or between the healthy and OA groups.

126



127

128 **Supplementary figure 1** Taxonomic identity of differentially present microbes across
129 conditions. Each row represents a species with significantly different abundances
130 (Mann–Whitney U-test FDR < 0.05) between the healthy and RA groups or between
131 the healthy and OA groups. Each column represents a sample from one of the groups
132 including the healthy, RAP1, RAP2, RAP3, RAP4, and OA groups. Color of each
133 heatmap unit represents the scaled abundance of a certain species in a specific sample.
134 Species are colored for significantly elevation (red) or depletion (green) in the arthritis
135 groups, in comparison with the healthy groups.

136 **Data accession**

137 Whole-genome shot-gun sequencing data are available in the Genome Sequence
138 Archive (GSA) section of National Genomics Data Center (project accession number
139 CRA004348) at <https://bigd.big.ac.cn/gsa/browse/CRA004348>.

140 **References**

- 141 1. Li D, Luo R, Liu CM, et al. MEGAHIT v1.0: A fast and scalable metagenome
142 assembler driven by advanced methodologies and community practices.
143 *Methods* 2016;102:3–11.
- 144 2. Hyatt D, Chen GL, Locascio PF, et al. Prodigal: prokaryotic gene recognition and
145 translation initiation site identification. *BMC Bioinformatics* 2010;11:119.
- 146 3. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of
147 protein or nucleotide sequences. *Bioinformatics* 2006;22:1658–9.
- 148 4. Pedersen HK, Forslund SK, Gudmundsdottir V, et al. A computational framework to
149 integrate high-throughput '-omics' datasets for the identification of potential
150 mechanistic links. *Nat Protoc* 2018;13:2781–800.
- 151 5. Nielsen HB, Almeida M, Juncker AS, et al. Identification and assembly of genomes
152 and genetic elements in complex metagenomic samples without using reference
153 genomes. *Nat Biotechnol* 2014;32:822–8.
- 154 6. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
155 2012;9(4):357–9.
- 156 7. Kanehisa M, Goto S, Sato Y, et al. Data, information, knowledge and principle: back
157 to metabolism in KEGG. *Nucleic Acids Res* 2014;42:D199–205.
- 158 8. Aramaki T, Blanc-Mathieu R, Endo H, et al. KofamKOALA: KEGG Ortholog
159 assignment based on profile HMM and adaptive score threshold. *Bioinformatics*
160 2020;36(7):2251–52.
- 161 9. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network
162 analysis. *BMC Bioinformatics* 2008;9:559.
- 163 10. Ceccarelli F, Massafra U, Perricone C, et al. Anti-TNF treatment response in
164 rheumatoid arthritis patients with moderate disease activity: a prospective
165 observational multicentre study (MODERATE). *Clin Exp Rheumatol*
166 2017;35(1):24–32.
- 167 11. Nakahara H, Nishimoto N. Anti-interleukin-6 receptor antibody therapy in
168 rheumatic diseases. *Endocr Metab Immune Disord Drug Targets*
169 2006;6(4):373–81.
- 170 12. Pedersen HK, Gudmundsdottir V, Nielsen HB, et al. Human gut microbes impact
171 host serum metabolome and insulin sensitivity. *Nature* 2016;535(7612):376–81.
- 172 13. Truong DT, Franzosa EA, Tickle TL, et al. MetaPhlAn2 for enhanced metagenomic
173 taxonomic profiling. *Nat Methods* 2015;12:902–3.

Supplementary information

A hidden link in gut-joint axis: Gut microbes promote rheumatoid arthritis at early stage by enhancing ascorbate degradation

Yan Zhao,^{1#} Mingyue Cheng,^{2#} Liang Zou,¹ Luxu Yin,¹ Chaofang Zhong,² Yugo Zha,² Xue Zhu,² Lei Zhang,^{3,4*} Kang Ning,^{2*} Jinxiang Han^{1*}

¹Shandong Medicine and Health Key Laboratory of Rheumatism, Shandong Key Laboratory of Rheumatic Disease and Translational Medicine, Department of Rheumatology and Autoimmunology, Shandong Provincial Qianfoshan Hospital, First Affiliated Hospital of Shandong First Medical University. Jinan 250014, Shandong, China

²Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Key Laboratory of Bioinformatics and Molecular-imaging, Center of AI Biology, Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

³Microbiome-X, National Institute of Health Data Science of China & Institute for Medical Dataology, Cheeloo College of Medicine, Shandong University, Jinan 250014, Shandong, China

⁴Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250014, Shandong, China

***Correspondence to** Dr Jinxiang Han, E-mail: jxhan@sdfmu.edu.cn; Dr Kang Ning, E-mail: ningkang@hust.edu.cn; Dr Lei Zhang, E-mail: zhanglei7@sdu.edu.cn.

[#]These authors contributed equally to this work

29 Sample description

30 A total of 122 fecal and 122 serum samples were collected from 122 outpatients from
31 the Shandong Provincial Qianfoshan Hospital (Jinan, Shandong, China). These
32 outpatients included 27 healthy individuals, 19 patients with osteoarthritis (OA), and
33 76 patients with rheumatoid arthritis (RA). Subsequently, the fecal samples were
34 sequenced and the serum samples were used to examine serum metabolites and
35 inflammatory cytokines. Serum inflammatory cytokines TNF- α and IL-6 were
36 quantified by the MESO SCALE DISCOVERY (MSD®) Quick Plex S600MM
37 multiplex assay. The cytokine levels of healthy individuals were extremely low and not
38 available. In addition, 95 knee-joint synovial fluid samples were collected from the RA
39 and OA patients to examine synovial fluid metabolites. Both serum and synovial fluid
40 metabolites were examined by UHPLC-MS/MS.

41 All of the participants were at fasting status during sample collection in the morning.

42 The participants were recruited in this study following the standards shown below:

- 43 1. Healthy individuals in good health condition with no gastrointestinal diseases, such
44 as diarrhea, constipation, and hematochezia, in the recent one month, no
45 hepatobiliary system diseases, no history of gastrointestinal tumors or inflammatory
46 diseases, no serious heart, liver, kidney, lung, brain or other organ disorders, no
47 infections, chronic diseases, or antibiotic treatment;
- 48 2. Healthy individuals had not taken any acid inhibitors, gastrointestinal motility drugs,
49 antibiotics, or living bacteria products such as yogurt in the recent one month;
- 50 3. Healthy individuals with no history or family history of mental illness, and no
51 history of gastrointestinal surgery;
- 52 4. RA/OA individuals with no other co-morbidity.

53 Metagenome sequencing and data processing

54 Whole-genome shot-gun sequencing of fecal samples were carried out on the Illumina
55 HiSeq X Ten. All samples were paired-end sequenced with a 150-bp read length. After

quality control, the paired-end reads were assembled into contigs using MEGAHIT (version 1.2.6)¹ with the minimum contig length set at 500 bp. The open reading frames (ORFs) were predicted from the assembled contigs using Prodigal (version 2.6.3)² with default parameters. The ORFs of <100 bp were removed. The ORFs were then clustered to remove redundancy using Cd-hit (version 4.6.6)³ with a sequence identity threshold set at 0.95 and the alignment coverage set at 0.9, which resulted in a catalog of 4,047,645 non-redundant genes. The non-redundant genes were then collapsed into metagenomic species (MGS)^{4,5} and grouped into KEGG functional modules.⁴

Identification of MGS

High-quality reads were mapped to the catalog of non-redundant genes using Bowtie 2 (version 2.2.9)⁶ with default parameters. The abundance profile for each catalogue gene was calculated as the sum of uniquely mapped sequence reads, using 19M sequence reads per sample (downsized). The co-abundance clustering of the 4,047,645 genes was performed using canopy algorithm (<http://git.dworzynski.eu/mgs-canopy-algorithm>),⁵ and 553 gene clusters that met the previously described criteria⁵ and contained more than 700 genes were referred to as MGS. MGS present in at least 4 samples were used for the following analysis. The abundance profiles of MGS were determined as the medium gene abundance throughout the samples. MGS were taxonomically annotated as described by Nielsen *et al.*⁵ and each MGS gene was annotated by sequence similarity to NCBI bacterial genome (BLASTN, E-value < 0.001)

Annotation of KEGG modules

The catalog of the non-redundant genes was functionally annotated to KEGG database (release 94.0) by KofamKOALA (version 1.3.0).^{7,8} The produced KEGG Orthologies (KOs) were mapped to the KEGG modules annotation downloaded on August 1, 2020 from the KEGG BRITE database. KOs present in at least 4 samples were used for the following analysis. The KO abundance profile was calculated by summing the

82 abundances of genes that were annotated to each KO.

83 **Clustering of co-abundant metabolites**

84 Co-abundant metabolites in serum or synovial fluid were identified using the R package
85 WGCNA⁹. As recommended by Pedersen *et al.*,⁴ a signed network and biweighted mid-
86 correlation were used for clustering with the soft threshold $\beta = 8$ for both serum and
87 synovial fluid metabolites. The minimum cluster size was set as 3. Similar clusters were
88 subsequently merged if the biweight mid-correlation between the cluster's eigen
89 vectors exceeded 0.8 for both serum and synovial fluid metabolites. The kIN of a
90 metabolite was calculated by summing connectivity with all other metabolites in the
91 given metabolite cluster. The kME was determined by the bicor-correlation between
92 the metabolite profile and module eigenvector. Both kIN and kME were used to
93 measure the intramodular hub-metabolite status.

94 **Cross-domain association analyses**

95 The clinical phenotypes, including types of arthritis (Healthy = 0, OA = 1, RA = 2) and
96 the levels of pro-inflammatory cytokines TNF- α and IL-6, were used in the association
97 analysis. TNF- α and IL-6 were selected based on their potentials to act as the
98 therapeutic targets for RA treatment.^{10 11} The associations between clinical phenotypes
99 and KEGG modules/metabolites clusters were determined through evaluating if the
100 Spearman correlations of the phenotype with the abundances of KOs/metabolites in the
101 given KEGG module/metabolite clusters were significantly higher or lower (Mann-
102 Whitney U-test FDR < 0.1) than with the abundances of all other KOs/metabolites. The
103 phenotypes adjusted by age and gender were also tested. Moreover, the union set of the
104 significant associations between KEGG modules and phenotypes/phenotypes adjusted
105 by age and gender, and the intersect set of the significant associations between
106 metabolites clusters and phenotypes/phenotypes adjusted by age and gender, were used
107 for the following association analysis. The associations between metabolite clusters and

108 KEGG modules were determined through evaluating if the Spearman correlations of
109 the eigen vectors of the metabolite clusters with the abundances of KOs in the given
110 KEGG module were significantly higher or lower (Mann–Whitney U-test FDR < 0.1)
111 than with the abundances of all other KOs/metabolites.

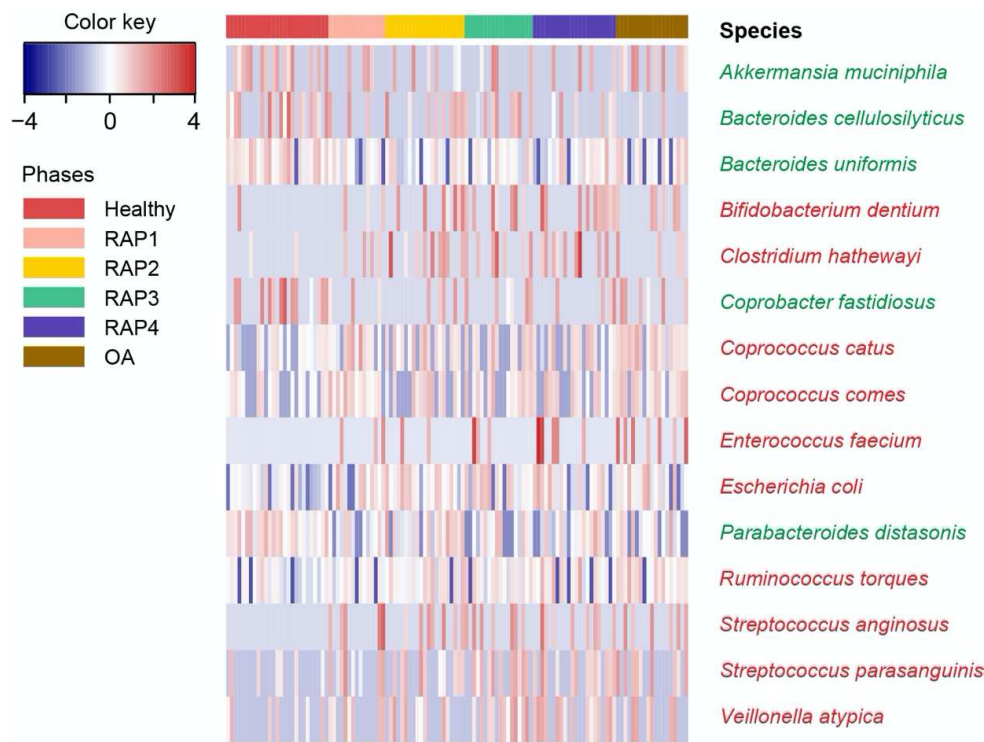
112 **Leave-one-out analysis**

113 Leave-one-out analysis was used to identify the specific MGS driving the observed
114 associations between KEGG module M00550 and the clinical phenotypes, including
115 the types of arthritis or the levels of pro-inflammatory cytokines TNF- α and IL-6. The
116 calculation of the KO abundance was iterated excluding the genes from a different MGS,
117 in each iteration. The effect of a given MGS on a specified association was defined as
118 the change in median Spearman correlation coefficient between KOs and clinical
119 phenotypes when genes from the respective MGS were left out, as previously
120 described.^{4 12}

121 **Taxonomic identity of differentially present microbes across conditions**

122 MetaPhlAn2¹³ was used to generate species profiles. Species that were present in less
123 than 10% samples were excluded. Supplementary Figure 1 displays the union set of the
124 species (n=15) with significantly different abundances (Mann–Whitney U-test FDR <
125 0.05) between the healthy and RA groups or between the healthy and OA groups.

126



127

128 **Supplementary figure 1** Taxonomic identity of differentially present microbes across
129 conditions. Each row represents a species with significantly different abundances
130 (Mann–Whitney U-test FDR < 0.05) between the healthy and RA groups or between
131 the healthy and OA groups. Each column represents a sample from one of the groups
132 including the healthy, RAP1, RAP2, RAP3, RAP4, and OA groups. Color of each
133 heatmap unit represents the scaled abundance of a certain species in a specific sample.
134 Species are colored for significantly elevation (red) or depletion (green) in the arthritis
135 groups, in comparison with the healthy groups.

136 **Data accession**

137 Whole-genome shot-gun sequencing data are available in the Genome Sequence
138 Archive (GSA) section of National Genomics Data Center (project accession number
139 CRA004348) at <https://bigd.big.ac.cn/gsa/browse/CRA004348>.

140 **References**

- 141 1. Li D, Luo R, Liu CM, et al. MEGAHIT v1.0: A fast and scalable metagenome
142 assembler driven by advanced methodologies and community practices.
143 *Methods* 2016;102:3–11.
- 144 2. Hyatt D, Chen GL, Locascio PF, et al. Prodigal: prokaryotic gene recognition and
145 translation initiation site identification. *BMC Bioinformatics* 2010;11:119.
- 146 3. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of
147 protein or nucleotide sequences. *Bioinformatics* 2006;22:1658–9.
- 148 4. Pedersen HK, Forslund SK, Gudmundsdottir V, et al. A computational framework to
149 integrate high-throughput '-omics' datasets for the identification of potential
150 mechanistic links. *Nat Protoc* 2018;13:2781–800.
- 151 5. Nielsen HB, Almeida M, Juncker AS, et al. Identification and assembly of genomes
152 and genetic elements in complex metagenomic samples without using reference
153 genomes. *Nat Biotechnol* 2014;32:822–8.
- 154 6. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
155 2012;9(4):357–9.
- 156 7. Kanehisa M, Goto S, Sato Y, et al. Data, information, knowledge and principle: back
157 to metabolism in KEGG. *Nucleic Acids Res* 2014;42:D199–205.
- 158 8. Aramaki T, Blanc-Mathieu R, Endo H, et al. KofamKOALA: KEGG Ortholog
159 assignment based on profile HMM and adaptive score threshold. *Bioinformatics*
160 2020;36(7):2251–52.
- 161 9. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network
162 analysis. *BMC Bioinformatics* 2008;9:559.
- 163 10. Ceccarelli F, Massafra U, Perricone C, et al. Anti-TNF treatment response in
164 rheumatoid arthritis patients with moderate disease activity: a prospective
165 observational multicentre study (MODERATE). *Clin Exp Rheumatol*
166 2017;35(1):24–32.
- 167 11. Nakahara H, Nishimoto N. Anti-interleukin-6 receptor antibody therapy in
168 rheumatic diseases. *Endocr Metab Immune Disord Drug Targets*
169 2006;6(4):373–81.
- 170 12. Pedersen HK, Gudmundsdottir V, Nielsen HB, et al. Human gut microbes impact
171 host serum metabolome and insulin sensitivity. *Nature* 2016;535(7612):376–81.
- 172 13. Truong DT, Franzosa EA, Tickle TL, et al. MetaPhlAn2 for enhanced metagenomic
173 taxonomic profiling. *Nat Methods* 2015;12:902–3.