Microbial dysbiosis and metabolic disorders promote rheumatoid arthritis progression across successive stages

Mingyue Cheng[1,2†], Yan Zhao[1†], Yazhou Cui[1], Yuguo Zha[2], Shufeng Li[1], Guangxiang Cao[1], Mian Li[1], Lei Zhang[3*], Kang Ning[2*], Jinxiang Han[1*]

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

[2]Key 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, China

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

[†]**These authors contributed equally to this work.**

[*]**Correspondence:**

Jinxiang Han (jxhan@sdfmu.edu.cn)

Kang Ning (ningkang@hust.edu.cn)

Lei Zhang (zhanglei7@sdu.edu.cn)

Keywords: Gut-joint axis, Rheumatoid arthritis, Microbiome, Metabolites, Dysbiosis, Pathogenesis, Successive stages, Osteoarthritis

Abstract

Rheumatoid arthritis (RA) is a chronic inflammatory disease that typically progresses through four stages, from mild stiffness to severe disability. The gut-joint axis is critical in RA pathogenesis, while its dynamic patterns and regulatory mechanisms throughout RA progression remain unknown. Here, we performed fecal metagenomic and serum metabolomic studies on 76 patients from four RA stages, 19 patients with osteoarthritis, and 27 healthy individuals. We demonstrated dynamics of microbial dysbiosis and metabolomic disorders and their crucial roles in RA pathogenesis across successive stages, which were in line with clinical RA progression. First, elevated *Collinsella aerofaciens* and depleted *Parabacteroides distasonis* increased gut permeability in the first stage. Second, the depletion of *Bacteroides uniformis* and *Bacteroides plebeius* led to decreased glycosaminoglycan metabolism, continuously hurting articular cartilage. Third, elevated *Escherichia coli* prompted an increase in ascorbate degradation particularly in the second and third stages, promoting inflammation and bone destruction. Fourth, abnormally high levels of serum methoxyacetic acid and cysteine-S-sulfate might inhibit osteoblasts in the second stage and enhance osteoclasts in the third stage, respectively, thus promoting bone erosion. Moreover, metagenomic and metabolomic features were identified for different RA stages, presenting profound diagnostic potential. Furthermore, combined with analysis of 16S rRNA genes and scanning electron microscopy on joint synovial fluid from

another cohort of 271 RA patients, microbial invasion of the joints such as *Eggerthella lenta* and *Bifidobacterium longum* was highly likely to happen in the fourth stage. Our findings elucidate the role of microbial dysbiosis and metabolic disorders in successive stages of RA, which open up new avenues for RA prognosis and therapy.

Introduction

Rheumatoid arthritis (RA) affects over tens of millions of people worldwide[1]. RA is recognized clinically as a chronic, inflammatory, autoimmune disease that primarily affects the joints and typically has four stages[2–5]: 1) In the first stage, the synovium of the joints is inflamed and most people have minor symptoms such as stiffness upon awakening; 2) In the second stage, the inflamed synovium has caused damage to the joint cartilage and people begin to feel swelling, and have a restricted range of motion; 3) In the third stage, RA has proceeded to a severe state when bone erosion begins and the cartilage on the surface of the bones has deteriorated, resulting in the bones rubbing against one another; and 4) In the fourth stage, certain joints are severely deformed and lose function and people suffer considerably from swelling, stiffness, mobility loss, and disability. To inhibit RA progression, specific therapeutic strategies are necessary for people across different RA stages.

Gut microbial dysbiosis has been implicated in the pathogenesis of RA via a range of mechanisms such as metabolic perturbation and immune response regulation, which is

64 known as the gut-joint axis[6–11]. For instance, increased abundance of *Prevotella* and
65 *Collinsella* in RA patients are correlated with the production of T_H17 cell cytokines[12,13].
66 Additionally, mice models have confirmed microbial roles in RA: Reduction of
67 *Bacteroides* spp. in mice with collagen-induced arthritis (CIA) may contribute to a local
68 pro-inflammatory environment by inhibiting CD4+ T cell differentiation into
69 regulatory T cells[14]. Microbial metabolites have also been correlated to immunity
70 regulation in RA: Short-chain fatty acids (SCFAs) have a local protective effect on
71 colonic inflammation, via binding to FFAR2 on colonic T cells and ILC3s[10,15]; SCFA
72 administration to mice with CIA can reduce the severity of arthritis by modulation of
73 IL-10[7,10]. Comprehensive metagenomic and metabolomic analyses could therefore
74 enhance our understanding about the gut-joint axis. However, the role of the gut-joint
75 axis across successive stages of RA is understudied[7,16], where more examination may
76 provide an alternative approach to ameliorate RA progression.

77 Here, we performed a multi-omics study including fecal metagenomics and serum
78 metabolomics on a cohort including 76 patients from four distinct RA stages, 19
79 patients with osteoarthritis (OA), and 27 healthy individuals. We depicted dynamic
80 patterns of microbial dysbiosis and metabolic disorders, and showed their continuous
81 roles throughout RA progression. We identified microbes related to the increased gut
82 permeability that occurred in the first stage, such as elevated *Collinsella aerofaciens*
83 and depleted *Parabacteroides distasonis*. We further identified microbes whose
84 depletion would continuously hurt articular cartilage through weakened

85 glycosaminoglycan metabolism, such as the *Bacteroides uniformis* and *Bacteroides*
86 *plebeius*. In addition, we found that most of microbial dysfunction driven by
87 *Escherichia coli* were correlated with induced inflammation and bone loss, particularly
88 in the second and third stage of RA. Moreover, we demonstrated that abnormal
89 elevations of two metabolites methoxyacetic acid and cysteine-S-sulfate inhibited
90 osteoblasts in the second stage and enhanced osteoclasts in the third stage, respectively.
91 These effects might partially explain why bone erosion clinically occurred in the third
92 stage of RA. Furthermore, together with analysis of 16S rRNA genes and scanning
93 electron microscopy on joint synovial fluid from another cohort of 271 RA patients, we
94 found that microbial invasion of the joints, such as *Eggerthella lenta* and
95 *Bifidobacterium longum*, might occur in the fourth stage of RA. Collectively, this study
96 highlighted that microbial dysbiosis and metabolic disorders promoted clinical
97 symptoms across successive stages of RA, which may open new opportunities for
98 prognosis, pain relief, and therapies throughout RA progression.

99 **Results**

100 **Comprehensive multi-omics characterization of RA**

101 We conducted a multi-omics study on gut metagenomic data and serum metabolomic
102 data from a cohort of 122 participants including 76 RA patients, 19 OA patients, and
103 27 healthy controls. To assess the gut-joint axis across successive stages of RA, RA
104 samples were grouped into four RA stages including RA stage 1 (RAS1, n = 15), RA

105 stage 2 (RAS2, n = 21), RA stage 3 (RAS3, n = 18), and RA stage 4 (RAS4, n = 22)
106 according to the rheumatoid diagnostic score[3] (**Supplementary Tables 1**). We
107 employed two parallel processes for gut metagenomic data analysis: 1) One was based
108 on 4 million predicted genes and investigated the functional composition across RA
109 stages and OA, as well as the metagenomic species (MGS) that most drove the
110 correlation of these microbial functions with RA or OA[17] (**Fig. 1a**); and 2) The other
111 reported the 232 classified microbial species composition across RA stages and OA[18],
112 profiled by MetaPhlAn2[19] (**Fig. 1b**). We also explored the variance in serum
113 metabolomic across RA stages and OA (**Fig. 1c**). Furthermore, we identified the most
114 prominent metagenomic and metabolomic features that could discriminate each of RA
115 stages or OA from the healthy state with high accuracy using the random forest
116 algorithm on microbial species, microbial KOs, and serum metabolites, respectively
117 (**Fig. 1d**). Finally, combined with analysis on joint synovial fluid from another cohort,
118 we attempted to test the existence of microbial invasion of the joints in RA progression.

119 **Dynamic patterns of microbial dysbiosis across successive stages of RA**

120 We firstly identified arthritis-related microbial species that were substantially elevated
121 or depleted in patients with RA or OA, when compared with the healthy controls ($P <$
122 0.05, FDR < 0.05, Mann-Whitney-Wilcoxon test, **Supplementary Tables 2**). The
123 union set of these arthritis-related species (n = 15) was then used for linear discriminant
124 analysis (LDA) to separate arthritis samples from healthy samples (**Fig. 2a**). We

125 discovered that arthritis gut microbiota could be discriminated from healthy microbiota,
126 while we did not find an evident separation between RA and OA samples in LDA. The
127 most abundant two arthritis-related species were *Bacteroides uniformis* whose
128 proportion was mainly contributed by healthy samples, and *Escherichia coli* whose
129 proportion was mainly contributed by RA and OA samples (**Fig. 2b**). Moreover, the
130 RA group contributed dominantly to the proportion of *Clostridium hathewayi*,
131 compared to its extremely low level in healthy and OA samples (**Fig. 2b,c)**.

132 Next, we sought to explore the taxonomic alterations across four successive stages of
133 RA. We examined 232 classified species and compared their abundances in each of RA
134 stages and OA to those in the healthy state (**Supplementary Tables 3**). The elevated
135 species in RA progression were mostly from the phyla Firmicutes and Actinobacteria,
136 while the depleted species were predominantly from the phylum Bacteroides (**Fig. 2d**).
137 Additionally, we exhibited the stage-specific abundance of the identified 15 arthritis-
138 related species, as well as species who or whose genera were previously recognized as
139 arthritis-related in other studies[20] (**Fig. 2d**).

140 For the arthritis-depleted species (**Fig. 2d**), we found that *B. uniformis* (RAS1: $P = 1.4$
141 $\times 10^{-3}$, RAS2: $P = 5.0 \times 10^{-3}$, RAS3: $P = 7.2 \times 10^{-3}$, RAS4: $P = 5.7 \times 10^{-3}$) and *C.*
142 *fastidiosus* (RAS1: $P = 0.041$, RAS2: $P = 0.024$, RAS3: $P = 0.022$) were significantly
143 depleted only during the RA progression but not in OA, whereas *Bacteroides*

144 *cellulosilyticus* was significantly depleted in OA ($P = 4.72 \times 10^{-4}$) and in the very early
145 and late stages of the RA progression (RAS1: $P = 0.023$, RAS4: $P = 6.24 \times 10^{-3}$).

146 For the arthritis-elevated species (**Fig. 2d**), we found that *C. hathewayi* was not
147 significantly elevated in the very early RA stage, though maintained a significantly
148 higher level during the subsequent RA stages than that in the healthy state (RAS2: $P =$
149 2.52×10^{-3}, RAS3: $P = 5.92 \times 10^{-3}$, RAS4: $P = 2.96 \times 10^{-3}$). In addition,
150 *Bifidobacterium dentium* also maintained considerable increased abundance from
151 RAS2 to RAS4 (RAS2: $P = 7.16 \times 10^{-3}$, RAS3: $P = 3.70 \times 10^{-3}$, RAS4: $P = 9.15 \times 10^{-}$
152 ^{4}). *B. dentium* was reported to be associated with the development of dental caries and
153 periodontal disease, both of which were particularly prevalent in patients with RA[21,22].
154 *Streptococcus anginosus* and *Streptococcus parasanguinis* maintained significant
155 elevation ($P < 0.05$) throughout the RA progression (**Fig. 2d**), while a few of other
156 species from genus *Streptococcus* were elevated stage-specifically. *Streptococcus mitis*
157 ($P = 1.21 \times 10^{-3}$) and *Streptococcus sanguinis* ($P = 5.37 \times 10^{-4}$) were mainly elevated
158 in RAS1, while *Streptococcus salivarius* was mainly elevated in RAS3 ($P = 0.035$) and
159 RAS4 ($P = 0.022$). These results suggested that although a few microbes were identified
160 as correlated with arthritis, not all remained plentiful or reduced throughout the
161 progression.

162 **Certain microbial species present stage-specific alteration throughout RA**
163 **progression**

164 Next, we identified species that were elevated or depleted exclusively in one specific
165 RA stage when compared to the healthy state ($P < 0.05$, Mann-Whitney-Wilcoxon test,
166 **Supplementary Tables 4**). *Collinsella aerofaciens*, for example, was found elevated
167 exclusively in RAS1 ($P = 0.043$). *C. aerofaciens* was previously reported to generate
168 severe arthritis when inoculated into collagen-induced arthritis (CIA)-susceptible mice,
169 and an *in vitro* experiment showed that *C. aerofaciens* could increase gut permeability
170 and induce IL-17A expression, a key cytokine involved in RA pathogenesis[13].
171 Therefore, the significant elevation of *C. aerofaciens* might contribute to the early
172 breach in gut barrier integrity during RA progression[7]. The translocation of microbial
173 products through the breach would then trigger the subsequent clinical arthritis[7].

174 Moreover, *Veillonella parvula*, whose infection could cause osteomyelitis[23], was found
175 elevated exclusively in RAS3 ($P = 0.027$). *Eggerthella lenta* ($P = 0.018$) and
176 *Bifidobacterium longum* ($P = 0.022$) were found elevated exclusively in RAS4.
177 Additionally, mice gavaged with *E. lenta* were reported to show an increased gut
178 permeability and produce proinflammatory cytokines IL-6, IL-21, IL-23[24]. We also
179 found that the species *Bacteroides ovatus*, *Bacteroides coprocola*, and *Bacteroides*
180 *faecis* were depleted exclusively in RAS1 ($P = 4.2 \times 10^{-3}$), RAS2 ($P = 0.036$), and
181 RAS4 ($P = 0.020$), respectively. *Dialister invisus* was found elevated exclusively in OA

182 microbiota (P = 0.041). *D. invisus* was previously reported to be more abundant in the
183 inflamed ileal and colonic biopsy from patients with spondyloarthritis than in the
184 noninflamed biopsy, and was positively correlated with spondyloarthritis severity[25].
185 These results suggested that certain microbial species in RA dysbiosis presented
186 elevated or depleted exclusively in a particular stage, for implementing specific
187 functions such as increasing gut permeability and inducing proinflammatory cytokines.

188 ***B. uniformis*- and *B. plebeius*-driven microbial dysfunction in glycosaminoglycan**
189 **metabolism continuously hurts articular cartilage**

190 To detect the microbial dysfunction in arthritis including RA and OA, we grouped
191 4,047,645 metagenomic genes into 6224 KEGG orthologs (KOs) and 404 KEGG
192 modules. We identified 12 KEGG modules that were significantly correlated with RA
193 or OA ($P < 0.05$, FDR < 0.1, Mann-Whitney-Wilcoxon test, **Fig. 3a**, **Extended Data**
194 **Fig. 1**, **Supplementary Tables 5** and **6**). We then compared KOs' abundances within
195 these arthritis-related KEGG modules between each of RA stages or OA with healthy
196 state (**Fig. 3a** and **Supplementary Tables 7**), and used leave-one-out analysis to
197 identify the metagenomic species (MGS) that most drove the correlations of these
198 KEGG modules with arthritis (**Fig. 3b** and **Extended Data Fig. 2**).

199 We found an evident decrease in glycosaminoglycan (CAG) metabolism across four
200 RA stages and OA. It was mainly reflected by the significant decrease in K01197
201 (hyaluronoglucosaminidase, RAS1: $P = 2.31 \times 10^{-3}$, RAS2: $P = 7.78 \times 10^{-5}$, RAS3: P

202 = 2.00 × 10^{-5}, RAS4: P = 6.70 × 10^{-4}, OA: 8.65 × 10^{-4}) of dermatan sulfate (DS)
203 degradation and the significant decrease in K10532 (heparan-alpha-glucosaminide N-
204 acetyltransferase, RAS1: P = 2.30 × 10^{-4}, RAS2: P = 0.045, RAS3: P = 3.2 × 10^{-3},
205 RAS4: P = 4.66 × 10^{-3}, OA: P = 3.72 × 10^{-4}) of heparan sulfate (HS) degradation (**Fig.**
206 **3a**). The compressive features of the articular cartilage are determined by CAGs,
207 predominantly chondroitin sulfate (CS) and hyaluronic acid[26]. CS is a major component
208 of the extracellular matrix of many connective tissues, including cartilage, bone, skin,
209 ligaments, and tendons[27]. We found that the significant depletion of DS degradation at
210 all the arthritis stages would decrease the production of chondroitin 4-sulfate (**Fig. 3b**),
211 which hurt the mechanical properties of articular cartilage of patients with arthritis.
212 Additionally, total CAGs, CS, and heparan sulfate/heparin (HS/H) were reported to be
213 higher in plasma of RA patients than in plasma of healthy subjects[28]. Moreover, the
214 sulfation of HS was reported to be increased in cartilages of patients with OA[29], and a
215 decreased HS content or a reduced sulfation level was reported to protect against OA
216 progression by regulating protease activity[26]. Thus, the significant depletion of HS
217 degradation in microbial function would prevent HS from being degraded, promoting
218 arthritis progression (**Fig. 3b**). The most driving species of dermatan sulfate
219 degradation and heparan sulfate degradation were MGS *Bacteroides uniformis* and
220 MGS *Bacteroides plebeius*, respectively. The genes of MGS *B. uniformis* related to
221 K01197 were found most depleted in RAS2, while the genes of MGS *B. plebeius* related
222 to K10532 were found most depleted in RAS3 and RAS4 (**Fig. 3B**). These results

223 indicated that the depleted microbial function in DS degradation and HS degradation
224 driven by *B. uniformis* and *B. plebeius*, respectively, could promote RA and OA in a
225 way of hurting articular cartilage.

226 **_E. coli_-driven microbial dysfunction promotes inflammation and bone lose**
227 **particularly in the second and third stages of RA**

228 We also identified certain arthritis-elevated microbial functions that were related to pro-
229 inflammation. As we have previously reported, *E. coli* could promote arthritis
230 progression by enhancing ascorbate degradation[11]. Ascorbate could prevent the
231 development of inflammatory arthritis by facilitating collagen synthesis, moderating
232 autoimmune responses, and ameliorating inflammation[30,31]. Here, we found most of the
233 KOs related to ascorbate degradation retained a higher level across RA stages and OA,
234 especially in RAS2 and RAS3 (**Fig. 3a**). Genes of K02821 (phosphotransferase system)
235 in RAS1, K03475 (phosphotransferase system), K03476 (L-ascorbate 6-phosphate
236 lactonase), and K03479 (L-ribulose-5-phosphate 3-epimerase) were mostly contributed
237 by MGS *E. coli* (**Fig. 3b**). Deficiency of the ascorbate reported in patients with RA[32]
238 might therefore be related to the enhanced ascorbate degradation in gut microbial
239 dysfunction. Moreover, we found that the enhanced ascorbate degradation was
240 evidently correlated with the elevation of multi serum cytokines, including IL-1β (P =
241 5.44×10^{-4}, q = 0.037), IL-17 ($P = 6.59 \times 10^{-4}$, q = 0.050), IL-8 ($P = 7.45 \times 10^{-4}$, q =
242 0.07), IL-2 ($P = 9.37 \times 10^{-4}$, q = 0.064), IL-6 ($P = 1.12 \times 10^{-3}$, q = 0.061), IL-12p70 (P

243 = 3.47 × 10^{-3}, q = 0.093), IL-4 (P = 4.2 × 10^{-3}, q = 0.097), TNF-α (P = 6.59 × 10^{-3}, q
244 = 0.050).

245 Moreover, to confirm the effects of ascorbate on RA progression, we examined the
246 serum TNF-α level and IL-6 level, bone CT scans, and bone density of 1) normal
247 DBA/1 mice, 2) DBA/1 mice with CIA, and 3) DBA/1 mice with CIA and gavage of
248 ascorbate. We found that the three-month gavage of ascorbate can prevent the increase
249 of TNF-α and IL-6 levels, inhibit bone destruction and bone loss (**Extended Data Fig.**
250 **3**).

251 For other arthritis-elevated microbial functions, the arginine succinyltransferase
252 pathway was found significantly elevated mainly in RAS2 and RAS3 (**Fig. 3a**). L-
253 arginine has been reported to be able to prevent bone loss induced by zinc oxide
254 nanoparticles or by cyclosporin A, through anti-inflammatory mechanism[33] or nitric
255 oxide production, respectively[34]. The most driven MGS of the arginine
256 succinyltransferase pathway was MGS *E. coli* (**Fig. 3b**). Moreover, another KEGG
257 module named trans-cinnamate degradation was also most driven by MGS *E. coli*,
258 where most KOs were significantly elevated in RA and OA, especially in RAS2 (**Fig.**
259 **3a,b**). Trans-cinnamate degradation was positively correlated with the elevation of
260 multi serum cytokines, including IL13 (P = 1.63 × 10^{-5}, q = 4.41 × 10^{-3}), IL-1β (P =
261 2.87 × 10^{-5}, q = 3.90 × 10^{-3}), IL10 (P = 4.10 × 10^{-5}, q = 5.57 × 10^{-3}), TNF-α (P = 6.97
262 × 10^{-5}, q = 9.47 × 10^{-3}), IL2 (P = 7.40 × 10^{-5}, q = 0.010), IL-8 (P = 2.46 × 10^{-4}, q =

263 0.067), IFNγ ($P = 3.92 \times 10^{-4}$, q = 0.036), IL4 ($P = 1.78 \times 10^{-3}$, q = 0.061), and IL-
264 12p70 ($P = 2.61 \times 10^{-3}$, q = 0.089). Furthermore, elevation of these microbial functions
265 were not as prominent in OA as they were in a particular RA stage (**Fig. 3a**), which
266 might help explain why OA had less inflammation than RA. Taken together, these
267 results suggested that arthritis-related microbial dysfunction could promote RA
268 progression by hurting bone tissue and strengthening inflammation. The inflammation-
269 related microbial dysfunction was extremely active in RAS2 and RAS3 and largely
270 driven by *E. coli*.

271 **Gut microbes may invade joint synovial fluid in the fourth stage of RA**

272 To examine whether and in which of stages microbes invaded joint synovial fluid of
273 patients with RA, we collected samples of joint synovial fluid from another cohort of
274 271 patients in four distinct stages of RA, including RAS1 (n = 52), RAS2 (n= 66),
275 RAS3 (n = 67), and RAS4 (n = 86). Enhanced gut permeability may render it possible
276 for microbes and their products to translocate, triggering an immune response[7,35].
277 Additionally, a few studies have shown the presence of microbial DNA in the joint
278 SF[36–39]. We thus speculated that gut microbes might invade the joint synovial fluid of
279 patients with RA through the gut-joint axis. To test this speculation, we next perform
280 16S rRNA gene sequencing on the SF samples. Notably, we were not able to obtain
281 enough bacterial DNA for sequencing in samples of RAS1, RAS2, or RAS3, however,
282 we could identify many microbes in samples of joint synovial fluid from patients in

283 RAS4 (**Extended Data Fig. 4**). We found that most of the microbes in joint synovial
284 fluid were from phyla Proteobacteria and Firmicutes (**Extended Data Fig. 4a**).
285 Moreover, we could recognize *Eggerthella lenta* and *Bifidobacterium longum* in most
286 of the synovial fluid samples (**Extended Data Fig. 4b**), both of which were observed
287 to be exclusively elevated in the intestine of patients in RAS4 (**Supplementary Tables**
288 **4**). In addition, *Prevotella copri* that has been reported highly correlated with RA[12,40]
289 was also found abundant in most synovial fluid samples of patients in RAS4.

290 We next randomly selected six synovial fluid samples per RA stage for bacteria
291 isolation. Only from three synovial fluid samples of RAS4 can we separate bacteria
292 (**Extended Data Fig. 5**). We then picked and sequenced three single colonies per
293 synovial fluid sample. Five of the nine colonies were identified as *Clostridium sp.*, three
294 were identified as *Enterococcus sp.*, and one was identified as *Citrobacter sp.*
295 (**Supplementary Tables 8**). We subsequently observed the corresponding synovial
296 fluid samples using scanning electron microscopy, and found substances shaped like
297 bacteria in rod-like or spherical forms (**Extended Data Fig. 6**). Taken together, this
298 multi-faceted investigation has provided unprecedented evidence to support the
299 existence of microbial invasion of the joints in the fourth stage of RA.

300 **Diagnostic potential of metagenomic and metabolomic features for different RA**
301 **stages**

302 Microbial dysbiosis has exhibited stage-specific patterns, playing a crucial role in
303 successive stages of RA, via microbial dysfunction. It was also speculated that
304 dysbiosis might promote RA progression via metabolic perturbation[7].

305 To determine the most discriminant characteristics of each RA stage and OA from
306 multi-omics perspective, we performed a random forest algorithm on 232 microbiome
307 species, 6224 KOs, and 277 metabolites (**Fig. 4**), and displayed the most notable
308 metagenomic and metabolomic features in **Fig. 5a**. Metabolites exhibited the best
309 AUROC in discriminating samples of four RA stages or OA from healthy samples, with
310 AUROC ranging from 0.974 to 0.998. Other characteristics at the species and KO levels
311 exhibited weaker discriminant ability than features at the metabolite level, with
312 AUROC ranging from 0.760 to 0.838 and from 0.799 to 0.852, respectively (**Fig. 4**).
313 Thus, metabolic disorders were predominant in each stage of RA, implicating a critical
314 involvement in pathogenesis and a precise diagnostic potential throughout the
315 development of RA.

316 **Dynamic patterns of metabolic disorders across successive stages of RA**

317 Each of the RA stages had its respective most prominent alteration in metabolites (**Fig.**
318 **4** and **Fig. 5a**), even though all of them were profoundly different from those of health
319 states (**Fig. 5b**). The most prominent changes in metabolites were the significant

320 increase of DL-lactate and gly-glu in RAS1 ($P = 2.15 \times 10^{-6}$, $P = 4.70 \times 10^{-4}$), the
321 decrease of N,N-dimethylaniline and the increase of methoxyacetic acid in RAS2 ($P =$
322 4.60×10^{-8}, $P = 1.28 \times 10^{-8}$), the increase of cysteine-S-sulfate ($P = 4.66 \times 10^{-12}$) in
323 RAS3, the increase of galactinol and 3α-mannobiose in RAS4 ($P = 5.71 \times 10^{-5}$, $P =$
324 5.68×10^{-4}), and the decrease of N,N-dimethylaniline and increase of gly-glu ($P = 2.00$
325 $\times 10^{-7}$, $P = 1.74 \times 10^{-3}$) in OA, as compared to a healthy state.

326 **Abnormal elevation of serum lactate occurs in the early stage of RA and is**
327 **correlated with inflammation**

328 A considerable increase in lactate, which was the most noticeable alteration in
329 metabolites in RAS1 (**Fig. 4a**), occurred early in RA development and remained
330 throughout the subsequent RA stages and OA (**Fig. 5b**). Lactate in the inflamed tissue
331 was reported to promote the upregulation of the lactate transporter SLC5A12 by human
332 CD4+ T cells[41]. This upregulation led to increased IL17 production via nuclear
333 PKM2/STAT3 and enhanced fatty acid synthesis[41]. In addition, antibody-mediated
334 blockade of SLC5A12 could ameliorate the severity of arthritis in a mouse model[41]. We
335 found serum DL-lactate was positively correlated with a certain number of
336 inflammatory cytokines such as IL-6 ($r = 0.33$, $P = 1.86 \times 10^{-5}$, $q = 3.69 \times 10^{-4}$), IL-17
337 ($r = 0.32$, $P = 4.23 \times 10^{-5}$, $q = 9.76 \times 10^{-4}$) and TNF-alpha ($r = 0.40$, $P = 1.84 \times 10^{-7}$, q
338 $= 1.28 \times 10^{-5}$). As the end product of glycolysis[42], the lactate production would be
339 affected by the intervention of glycolysis, such as the disruption of the glucose-6-

340 phosphate isomerase (GPI, K06859) and phosphofructokinase (PFK, K00850). Both
341 GPI and PFK, as the upstream glycolytic enzyme, were found significantly decreased
342 in RA stages ($P < 0.05$, $q < 0.1$), while GPI was not significantly decreased in the OA
343 stage (**Fig. 3a**). As a consequence, microbial functional variation in glycolysis might
344 result in an increased concentration of DL-lactate across RA stages, contributing to the
345 elevated inflammatory cytokines.

346 **Metabolic disorders break the balance between osteoblasts and osteoclasts mainly**
347 **in the second and third stages of RA**

348 Furthermore, we identified that a few noticeable alterations in metabolites might
349 distinguish a given RA stage from not just healthy controls but also other RA stages or
350 OA. (**Fig. 5b**): Methoxyacetic acid in RAS2 ($P = 1.68 \times 10^{-4}$) or cysteine-S-sulfate in
351 RAS3 ($P = 2.42 \times 10^{-4}$) or Galactinol and 3α-mannobiose in RAS4 ($P = 9.37 \times 10^{-3}$, P
352 $= 4.89 \times 10^{-3}$), respectively, was higher than that in all the other RA stages and OA. In
353 addition, DL-lactate in OA was less than that in all RA stages ($P = 0.037$). These
354 metabolic distinctions may improve clinical differentiation of early RA from OA.
355 Such dramatic changes in serum metabolites suggested that they might be more than
356 just a metabolomic depiction of each RA stage; they could also be potential promoters
357 for the next stage. For instance, methoxyacetic acid was reported to have inhibitory
358 effects on osteoblasts and could cause reductions in bone marrow cellularity[43–45]. The
359 dominant elevation of methoxyacetic acid in RAS2 might be a precursor to bone erosion

360 that occurred in RAS3. Moreover, the notable elevation of cysteine-S-sulfate in RAS3
361 could promote the osteoclasts. Cysteine-S-sulfate, a structural analog of glutamate,
362 acted as an agonist of N-methyl-D-aspartate receptor (NMDA-R) whose expression and
363 function in osteoclasts engaged in bone resorption[46]. Therefore, aberrant elevations of
364 methoxyacetic acid in RAS2 might hinder osteoblasts, whereas abnormal elevations of
365 cysteine-S-sulfate in RAS3 might encourage osteoclasts. The imbalance between
366 osteoblasts and osteoclasts promoted the bone erosion that occurred clinically in the
367 third stage of RA.

368 **Discussion**

369 We performed metagenomic and metabolomic studies on a cohort of 76 RA patients,
370 19 OA patients, and 27 healthy controls, together with 16S rRNA gene sequencing and
371 scanning electron microscopy on joint synovial fluid samples from another cohort of
372 271 RA patients, and drew several important conclusions. First, we found that microbial
373 dysbiosis and metabolic disorders presented dynamic patterns across successive stages
374 of RA. Second, these dysbiosis and disorders continuously promoted RA progression
375 by increasing gut permeability, inducing inflammation, and hurting bone tissue. In
376 particular, microbial dysfunction in dermatan sulfate and heparan sulfate degradation
377 continuously hurts the articular cartilage. Third, certain influences were more focused
378 on a particular stage of RA. Notably, microbial dysfunction related to inflammation
379 was focused on the second and third stages and was mostly driven by *E. coli*. Metabolic

380 disorders prompted bone erosion through the abnormal elevation of methoxyacetic acid
381 in the second stage and cysteine-S-sulfate in the third stage. Finally, there was strong
382 evidence that microbial invasion of the joints happened in the fourth stage of RA.
383 Synovial fluid 16S rRNA gene sequencing was capable of detecting *Eggerthella lenta*
384 and *Bifidobacterium longum*, both of which were elevated in fecal metagenomics of
385 patients, exclusively in the fourth stage of RA rather than other stages. Moreover,
386 *Clostridium sp.*, *Enterococcus sp.*, and *Citrobacter sp.* were separated from the joint
387 synovial fluid. Under scanning electron microscopy, we could observe substances
388 shaped like bacteria in rod-like or spherical forms. However, we were unable to detect
389 bacteria in the joint synovial fluid from patients in other stages of RA due to the low
390 DNA content. We conclude that the buildup of microbial dysbiosis and metabolic
391 disorders across successive RA stages may induce the microbial invasion of the joints
392 in the last stage of RA.

393 The findings of the continuous effects of microbial dysbiosis and metabolic disorders
394 across successive RA stages could be of considerable clinical importance (**Fig. 6**). First,
395 hurting bone tissue (**Fig. 6a**). Depleted *B. uniformis* drove the decreased dermatan
396 sulfate degradation, leading to less production of chondroitin 4-sulfate. The low content
397 of chondroitin 4-sulfate would inhibit the formation of articular cartilage[26,27].
398 Additionally, depleted *B. plebeius* drove the less degradation of HS. A high HS content
399 would promote OA progression[26,29]. Second, increasing inflammation particularly in
400 the second and the third of RA (**Fig. 6b**). Elevated *E. coli* drove enhanced degradation

401 of L-Arginine through the arginine succinyltransferanse pathway. L-Arginine acted as
402 an inhibitor of bone loss via anti-inflammation[33] and nitric oxide production[34]. Besides,
403 elevated *E. coli* drove enhanced ascorbate degradation. Ascorbate is an anti-
404 inflammatory vitamin[30,31]. We found that three months of ascorbate treatment to mice
405 with CIA could relieve arthritis, by preventing bone destruction and inflammation.
406 Owing to the sequencing limitation in this study, we are unable to pinpoint the strain of
407 microbial species in these gut-joint axes. Further explorations are required for
408 determining strain identity, maybe through third-generation sequencing, as well as for
409 testing its effects in a sterile mice model. In clinical practice, the intervention based on
410 these proposed mechanisms may profoundly improve the therapy for patients during
411 the whole RA progression, through restoring *B. uniformis* and *B. plebeius*, and
412 inhibiting *E. coli*, or supplying patients with ascorbate and chondroitin 4-sulfate.

413 Apart from clinical treatments for microbial dysbiosis and metabolic disorders in the
414 whole RA progression, specific treatments are necessary for each RA stage to not only
415 relieve the symptoms in the current stage but also inhibit the onset of the subsequent
416 stages. Our findings suggested that certain effects of microbial dysbiosis or metabolic
417 disorders on RA progression were more focused on a particular stage. In the early stage
418 of RA, elevated *C. aerofaciens* and depleted *Parabacteroides. Distasonis* both
419 contributed to the increased gut permeability (**Fig. 6b**). *C. aerofaciens* was found
420 elevated exclusively in the first stage of RA, rather than other stages. *C. aerofaciens*
421 was reported to induce expression of IL-17 network cytokines and increase gut

422 permeability via reducing the expression of tight junction protein in epithelial cells[13].
423 *P. distasonis* was depleted in most of RA stages and OA with a remarkable ability to
424 distinguish arthritis samples from healthy samples, particularly in the first stage (**Fig.**
425 **5b**). Previous studies have reported the anti-inflammation effect of *P. distasonis* such
426 as its correlation with CRP[47], TNF-α[48], and IL-1β[48], and that the oral administration of
427 its antigens could reduce intestinal inflammation in colitis mice[49]. However, the role of
428 *P. distasonis* in RA pathogenesis remained unknown. A recent study found that *P.*
429 *distasonis* could preserve the intestinal barrier integrity in azoxymethane-treated A/J
430 mice, and possess anti-inflammatory properties and anti-tumor properties through the
431 inhibition of signaling via TLR4, MYD88, and Akt, as well as the stimulation of
432 apoptosis[50]. We thus speculate that the considerable depletion of *P. distasonis* occurs
433 in the very early of RA, contributing to inflammation and the collapse of intestinal
434 barrier integrity, though this possible mechanism warrants further investigations.
435 Intervention on *C. aerofaciens* and *P. distasonis* therefore may have a higher priority
436 in the first stage of RA than in subsequent stages to inhibit the early increase of gut
437 permeability and inflammation.

438 In the second stage of RA, the aberrant elevation of methoxyacetic acid had inhibitory
439 effects on osteoblasts and caused reductions in bone marrow cellularity[43–45] (**Fig. 6b**).
440 The inhibited osteoblasts drew the foreshadowing for the bone erosion that happened
441 in the next stage. In the third stage, the considerable elevation of cysteine-S-sulfate
442 might enhance the osteoclasts by NMDA-R interaction[46] (**Fig. 6b**). The imbalance

443 between osteoblasts and osteoclasts then promoted bone erosion that happened in the
444 third stage and persisted in the late RA stages (**Fig. 6b**). Thus, methoxyacetic acid may
445 be a targeted metabolite for treatment to patients in the second stage of RA and serve
446 as a precaution against the upcoming third stage. In the fourth stage of RA, continuous
447 damages in gut permeability, intestinal immunity, and bone tissue resulted in the
448 extremely high elevation of proinflammatory cytokines and severe bone erosion, and
449 even microbial invasion of the joints. Due to the collapse of the intestinal barrier,
450 microbes and microbial metabolites might be transferred to the joints via blood[7]. In
451 addition to routine medical therapies for immunity regulation, specific treatments to the
452 microbes in the joint synovial fluid of the patients in the fourth stage of RA may
453 ameliorate the joint micro-environment to decrease synovial inflammation and prevent
454 potential bacterial effects.

455 This study also has prospects and limitations. Firstly, it may aid in enhancing RA
456 therapy across populations by assessing whether RA patients with different genetic
457 backgrounds exhibit different gut-joint links across stages. Secondly, a long-term
458 follow-up investigation on a single individual throughout his/her RA development may
459 reinforce the conclusions of this study. Thirdly, it remains unclear how bacterial genetic
460 materials are transferred from intestine to joint. It might be realized by bacteria
461 transmission through blood or by means of extracellular vesicles or both. Fourthly, the
462 proposed links between microbial dysbiosis/metabolic disorders and RA can serve as a
463 guidance for future studies on RA pathogenesis. Lastly, additional researches into the

464 synovial fluid microbiome and metabolome have the potential to figure out more
465 sophisticated mechanisms underlying RA pathogenesis.

466 In conclusion, we demonstrate dynamic shifts in gut microbiome and serum
467 metabolome, and their continuous roles in pathogenesis of RA across four successive
468 stages. Microbial dysbiosis and metabolic disorders promotes RA through increasing
469 gut permeability, inducing inflammation, and hurting bone tissue. Although these
470 effects persist in the whole RA progression, certain effects are more focused on specific
471 stages. Furthermore, the buildup of the effects of these dysbiosis and disorders may
472 lead to the collapse in the fourth stage of RA, as represented by the microbial invasion
473 of the joints. This study highlights that microbial dysbiosis and metabolic disorders
474 promote clinical symptoms across successive stages of RA. The stage-specific
475 intervention in microbial dysbiosis and metabolic disorders was warranted for
476 prognosis and prevention of RA progression.

477 **Main references**

478 1. Almutairi, K., Nossent, J., Preen, D., Keen, H. & Inderjeeth, C. The global
479 prevalence of rheumatoid arthritis: a meta-analysis based on a systematic review.
480 *Rheumatol. Int.* **41**, 863–877 (2021).

481 2. Smolen, J.S., Aletaha, D. & McInnes, I.B. Rheumatoid arthritis. *Lancet* **388**,
482 2023–2038 (2016).

483 3. Aletaha, D.*, et al.* 2010 Rheumatoid arthritis classification criteria: an American

484 College of Rheumatology/European League Against Rheumatism collaborative
485 initiative. *Arthritis Rheumatol* **62**, 2569–2581 (2010).
486 4. Coutant, F. & Miossec, P. Evolving concepts of the pathogenesis of rheumatoid
487 arthritis with focus on the early and late stages. *Curr. Opin. Rheumatol.* **32**, 57–
488 63 (2020).
489 5. Barhum, L. What Does Rheumatoid Arthritis Progression Look Like?
490 https://www.verywellhealth.com/rheumatoid-arthritis-stages-of-progression-
491 4768891 (2021).
492 6. Wells, P.M.*, et al.* Associations between gut microbiota and genetic risk for
493 rheumatoid arthritis in the absence of disease: a cross-sectional study. *Lancet*
494 *Rheumatol* **2**, e418–e427 (2020).
495 7. Zaiss, M.M., Joyce Wu, H.J., Mauro, D., Schett, G. & Ciccia, F. The gut-joint
496 axis in rheumatoid arthritis. *Nat. Rev. Rheumatol.* **17**, 224–237 (2021).
497 8. Jubair, W.K.*, et al.* Modulation of Inflammatory Arthritis in Mice by Gut
498 Microbiota Through Mucosal Inflammation and Autoantibody Generation.
499 *Arthritis Rheumatol* **70**, 1220–1233 (2018).
500 9. Takahashi, D.*, et al.* Microbiota-derived butyrate limits the autoimmune
501 response by promoting the differentiation of follicular regulatory T cells.
502 *EbioMedicine* **58**, 102913 (2020).
503 10. Smith, P.M.*, et al.* The microbial metabolites, short-chain fatty acids, regulate
504 colonic Treg cell homeostasis. *Science* **341**, 569–573 (2013).

505 11. Zhao, Y., *et al.* Hidden link in gut-joint axis: gut microbes promote rheumatoid
506 arthritis at early stage by enhancing ascorbate degradation. *Gut* (2021).
507 12. Pianta, A., *et al.* Evidence of the Immune Relevance of Prevotella copri, a Gut
508 Microbe, in Patients With Rheumatoid Arthritis. *Arthritis Rheumatol* **69**, 964–
509 975 (2017).
510 13. Chen, J., *et al.* An expansion of rare lineage intestinal microbes characterizes
511 rheumatoid arthritis. *Genome Med.* **8**, 43 (2016).
512 14. Round, J.L. & Mazmanian, S.K. Inducible Foxp3+ regulatory T-cell
513 development by a commensal bacterium of the intestinal microbiota. *Proc. Natl.*
514 *Acad. Sci. U. S. A.* **107**, 12204–12209 (2010).
515 15. Chun, E., *et al.* Metabolite-Sensing Receptor Ffar2 Regulates Colonic Group 3
516 Innate Lymphoid Cells and Gut Immunity. *Immunity* **51**, 871–884.e876 (2019).
517 16. Maeda, Y. & Takeda, K. Role of Gut Microbiota in Rheumatoid Arthritis. *J Clin*
518 *Med* **6**, 60 (2017).
519 17. Pedersen, H.K., *et al.* A computational framework to integrate high-throughput
520 '-omics' datasets for the identification of potential mechanistic links. *Nat Protoc*
521 **13**, 2781-2800 (2018).
522 18. Li, D., *et al.* MEGAHIT v1.0: A fast and scalable metagenome assembler driven
523 by advanced methodologies and community practices. *Methods* **102**, 3–11
524 (2016).
525 19. Truong, D.T., *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic

526 profiling. *Nat. Methods* **12**, 902–903 (2015).

527 20. de Oliveira, G.L.V., Leite, A.Z., Higuchi, B.S., Gonzaga, M.I. & Mariano, V.S.
528 Intestinal dysbiosis and probiotic applications in autoimmune diseases.
529 *Immunology* **152**, 1–12 (2017).

530 21. Lugli, G.A.*, et al.* Decoding the Genomic Variability among Members of the
531 Bifidobacterium dentium Species. *Microorganisms* **8**, 1720 (2020).

532 22. Silvestre-Rangil, J., Bagán, L., Silvestre, F.J. & Bagán, J.V. Oral manifestations
533 of rheumatoid arthritis. A cross-sectional study of 73 patients. *Clin Oral Investig*
534 **20**, 2575-2580 (2016).

535 23. Marriott, D., Stark, D. & Harkness, J. Veillonella parvula discitis and secondary
536 bacteremia: a rare infection complicating endoscopy and colonoscopy? *J. Clin.*
537 *Microbiol.* **45**, 672–674 (2007).

538 24. Balakrishnan, B., Luckey, D. & Taneja, V. Autoimmunity-Associated Gut
539 Commensals Modulate Gut Permeability and Immunity in Humanized Mice.
540 *Mil. Med.* **184**, 529–536 (2019).

541 25. Tito, R.Y.*, et al.* Brief Report: Dialister as a Microbial Marker of Disease
542 Activity in Spondyloarthritis. *Arthritis Rheumatol* **69**, 114–121 (2017).

543 26. Severmann, A.C.*, et al.* An altered heparan sulfate structure in the articular
544 cartilage protects against osteoarthritis. *Osteoarthr Carti* **28**, 977–987 (2020).

545 27. Henrotin, Y., Mathy, M., Sanchez, C. & Lambert, C. Chondroitin sulfate in the
546 treatment of osteoarthritis: from in vitro studies to clinical recommendations.

547 *Ther. Adv. Musculoskelet. Dis.* **2**, 335–348 (2010).

548 28. Jura-Półtorak, A.*, et al.* Alterations of plasma glycosaminoglycan profile in
549 patients with rheumatoid arthritis in relation to disease activity. *Clin. Chim. Acta*
550 **433**, 20–27 (2014).

551 29. Shamdani, S.*, et al.* Heparan sulfate functions are altered in the osteoarthritic
552 cartilage. *Arthritis Res. Ther.* **22**, 283 (2020).

553 30. Pattison, D.J.*, et al.* Vitamin C and the risk of developing inflammatory
554 polyarthritis: prospective nested case-control study. *Ann. Rheum. Dis.* **63**, 843-
555 847 (2004).

556 31. Carr, A.C. & Maggini, S. Vitamin C and Immune Function. *Nutrients* **9**(2017).

557 32. Abrams, E. & Sandson, J. EFFECT OF ASCORBIC ACID ON RHEUMATOID
558 SYNOVIAL FLUID. *Ann. Rheum. Dis.* **23**, 295–299 (1964).

559 33. Abdelkarem, H.M., Fadda, L.H., El-Sayed, E.M. & Radwan, O.K. Potential
560 Role of L-Arginine and Vitamin E Against Bone Loss Induced by Nano-Zinc
561 Oxide in Rats. *J. Diet. Suppl.* **15**, 300–310 (2018).

562 34. Fiore, C.E.*, et al.* L-arginine prevents bone loss and bone collagen breakdown
563 in cyclosporin A-treated rats. *Eur. J. Pharmacol.* **408**, 323–326 (2000).

564 35. Manfredo Vieira, S.*, et al.* Translocation of a gut pathobiont drives
565 autoimmunity in mice and humans. *Science* **359**, 1156–1161 (2018).

566 36. Mårdh, P.A., Nilsson, F.J. & Bjelle, A. Mycoplasmas and bacteria in synovial
567 fluid from patients with arthritis. *Ann. Rheum. Dis.* **32**, 319–325 (1973).

568 37. Zhao, Y., *et al.* Detection and characterization of bacterial nucleic acids in
569 culture-negative synovial tissue and fluid samples from rheumatoid arthritis or
570 osteoarthritis patients. *Sci. Rep.* **8**, 14305 (2018).
571 38. Siala, M., *et al.* Analysis of bacterial DNA in synovial tissue of Tunisian patients
572 with reactive and undifferentiated arthritis by broad-range PCR, cloning and
573 sequencing. *Arthritis Res. Ther.* **10**, R40 (2008).
574 39. Cox, C.J., Kempsell, K.E. & Gaston, J.S. Investigation of infectious agents
575 associated with arthritis by reverse transcription PCR of bacterial rRNA.
576 *Arthritis Res. Ther.* **5**, R1–8 (2003).
577 40. Scher, J.U., *et al.* Expansion of intestinal Prevotella copri correlates with
578 enhanced susceptibility to arthritis. *Elife* **2**, e01202 (2013).
579 41. Pucino, V., *et al.* Lactate Buildup at the Site of Chronic Inflammation Promotes
580 Disease by Inducing CD4(+) T Cell Metabolic Rewiring. *Cell Metab.* **30**, 1055–
581 1074 (2019).
582 42. Rogatzki, M.J., Ferguson, B.S., Goodwin, M.L. & Gladden, L.B. Lactate is
583 always the end product of glycolysis. *Front. Neurosci.* **9**, 22 (2015).
584 43. Brown, N.A., Holt, D. & Webb, M. The teratogenicity of methoxyacetic acid in
585 the rat. *Toxicol. Lett.* **22**, 93–100 (1984).
586 44. Miller, R.R., Carreon, R.E., Young, J.T. & McKenna, M.J. Toxicity of
587 methoxyacetic acid in rats. *Fundam Appl Toxicol* **2**, 158–160 (1982).
588 45. Sparks, N.R. *The Embryotoxic Effects of Harm Reduction Tobacco Products on*

589 *Osteoblasts Developing from Human Embryonic Stem Cell* (University of
590 California, Riverside, 2018).
591 46. Li, P.*, et al.* Metabolic Alterations in Older Women With Low Bone Mineral
592 Density Supplemented With Lactobacillus reuteri. *JBMR Plus* **5**, e10478 (2021).
593 47. Picchianti-Diamanti, A.*, et al.* Analysis of Gut Microbiota in Rheumatoid
594 Arthritis Patients: Disease-Related Dysbiosis and Modifications Induced by
595 Etanercept. *Int. J. Mol. Sci.* **19**, 2938 (2018).
596 48. Schirmer, M.*, et al.* Linking the Human Gut Microbiome to Inflammatory
597 Cytokine Production Capacity. *Cell* **167**, 1125–1136.e1128 (2016).
598 49. Kverka, M.*, et al.* Oral administration of Parabacteroides distasonis antigens
599 attenuates experimental murine colitis through modulation of immunity and
600 microbiota composition. *Clin Exp Immunol* **163**, 250–259 (2011).
601 50. Koh, G.Y., Kane, A.V., Wu, X. & Crott, J.W. Parabacteroides distasonis
602 attenuates tumorigenesis, modulates inflammatory markers and promotes
603 intestinal barrier integrity in azoxymethane-treated A/J mice. *Carcinogenesis* **41**,
604 909–917 (2020).

605 **Figure legends**

606 **Fig. 1 | The analysis pipeline for the multi-omics study on patients with RA.** Gut
607 microbiome whole-genome sequencing (WGS) data and serum metabolomic data from
608 76 rheumatoid arthritis (RA), 19 osteoarthritis (OA) patients, and 27 healthy individuals
609 were used in the analysis. RA samples were divided into four stages including RAS1
610 (n = 15), RAS2 (n = 21), RAS3 (n = 18), RAS4 (n = 22) according to the rheumatoid
611 diagnostic score. **a**, Microbial genes, predicted from WGS sequences, were clustered
612 into metagenomic species (MGS) for taxonomic annotation and mapped to KEGG
613 orthology for functional annotation. Ten KEGG modules that were significantly
614 correlated with arthritis (Healthy = 0, OA = 1, RA = 2, $P < 0.05$, FDR < 0.05, Mann-
615 Whitney-Wilcoxon test) were remained to describe the stage-specific functional pattern.
616 Additionally, the leave-one-out analysis was used to determine the MGS that most
617 drove the correlation of these ten KEGG modules with arthritis (Methods). KEGG maps
618 of these specific modules, their KOs' abundances, and the corresponding driving MGS
619 abundances were plotted to show the stage-specific functional driving genes variation.
620 **b**, A total of 232 classified microbial species, identified from WGS sequences, were
621 used to describe the stage-specific taxonomic pattern. **c**, In total, 277 serum metabolites
622 were used to describe the stage-specific metabolomic pattern. **d**, Random forest
623 algorithm was performed on 6224 KOs, 232 microbial species, and 277 serum
624 metabolites from (a–c) to detect the stage-specific biomarkers at the multi-omics level
625 and the AUROC in classifying stage-specific arthritis samples from healthy individuals.

626

627 **Fig. 2 | Microbial dysbiosis across four stages of RA. a**, Linear discriminant analysis
628 to separate the samples of healthy, OA, and RA groups. The length and direction of the
629 arrows indicate the normalized scalings for each of the microbial species. **b**, Each row
630 represents a species with significantly different abundances ($P < 0.05$, FDR < 0.05)
631 between the RA and the healthy groups or between the OA and the healthy groups.
632 Each column represents a sample from one of the groups including the healthy (n = 27),
633 RAS1 (n = 15), RAS2 (n = 21), RAS3 (n = 18), RAS4 (n = 22), and OA (n = 19) groups.
634 The color of each heatmap unit represents the scaled abundance of a certain species in
635 a specific sample. Species are colored for significant elevation (red) or depletion (green)
636 in the arthritis groups, in comparison with the healthy groups. **c**, Three apices of the
637 triangle in the ternary diagram represent components of healthy, OA, and RA groups.
638 Microbial species in (b) are plotted as dots inside the triangle, the size of which indicates
639 the relative abundance. For each dot, moving away from the apex, the proportion of the
640 component at the apex decreases as the proportion of the remaining two increases. **d**,
641 In total, 232 classified species whose relative abundance was more than 1×10^{-6} are
642 shown in a phylogenetic tree, grouped in the phyla Fusobacteria, Proteobacteria,
643 Actinobacteria, Bacteroidetes, and Firmicutes. In the outer circles, species at each
644 arthritis stage are marked for significant elevation ($P < 0.05$, orange; FDR < 0.05, red)
645 or depletion ($P < 0.05$, light green; FDR < 0.05, dark green) in abundance, as compared
646 to those of healthy individuals. The innermost circle shows species relative abundances

averaged over all samples. Box plots show the relative abundances of species recorded in (b) and (c), and others that have been correlated with RA in previous work. The order and color of the box are consistent with the stage circle. Significant changes (elevation and depletion) are denoted as follows: +++, elevation with $P < 0.005$; ++, elevation with $P < 0.01$; +, elevation with $P < 0.05$; −−−, depletion with $P < 0.005$; −−, depletion with $P < 0.01$; −, depletion at $P < 0.05$. Boxes represent the interquartile range between the first and third quartiles and the line inside represents the median. Whiskers denote the lowest and highest values within the 1.5 × interquartile range from the first and third quartiles, respectively. All the statistical significance was calculated by the Mann-Whitney-Wilcoxon test. Microbial species shared in (b−d) are marked with consistent circular marks.

658

659 **Fig. 3 | Microbial dysfunction across four stages of RA.** Gene abundances were
660 assessed for elevation or depletion in each of the arthritis stages, RAS1 (n = 15), RAS2
661 (n = 21), RAS3 (n = 18), RAS4 (n = 22), and OA (n = 19) compared to the healthy
662 individuals (n = 27). **a**, Relative abundance of KO genes in the KEGG modules that
663 were significantly correlated with arthritis ($P < 0.05$, FDR < 0.05). KO genes with a
664 prevalence of 5% or higher are shown. Significant changes (elevation and depletion)
665 are denoted as follows: +++, elevation with $P < 0.005$; ++, elevation with $P < 0.01$; +,
666 elevation with $P < 0.05$; −−−, depletion with $P < 0.005$; −−, depletion with $P < 0.01$; −,
667 depletion at $P < 0.05$. **b**, KO genes involved in specific KEGG pathway modules in (a)
668 are shown in the KEGG pathway maps. Each box in a pathway represents a KO gene
669 and is marked in red for elevation or in blue for depletion at any of the stages compared
670 to healthy individuals. Bar plots show relative gene abundances averaged over samples
671 within each of the five groups (healthy (H), RAS1 (S1), RAS2 (S2), RAS3 (S3), RAS4
672 (S4), and OA) and are colored according to the values. Each KO gene is composed of
673 MGS genes represented by circles. The sizes and colors of the circles are proportional
674 to the relative abundances of the MGS genes. MGS genes are grouped into one row and
675 indicated by the taxonomic name. The three MGS that most drove the correlation of the
676 KEGG modules with arthritis types are shown. Significant changes (elevation and
677 depletion) are denoted as follows: +++, elevation with $P < 0.005$; ++, elevation with P
678 < 0.01; +, elevation with $P < 0.05$; −−−, depletion with $P < 0.005$; −−, depletion with

679 $P < 0.01$; −, depletion at $P < 0.05$. The boxes represent 25th–75th percentiles, black lines

680 indicate the median, and whiskers extend to the maximum and minimum values within

681 1.5 × the interquartile range. All the statistical significance was calculated by the Mann-

682 Whitney-Wilcoxon test.

683

684 **Fig. 4 | Multi-omics biomarkers and their diagnostic potential for discriminating**
685 **each of RA stages from healthy state.** A random forest algorithm was performed on
686 6224 KOs, 232 microbial species, and 277 serum metabolites to detect the stage-
687 specific biomarkers in the arthritis progression including RAS1 (**a**), RAS2 (**b**), RAS3
688 (**c**), RAS4 (**d**), and OA (**e**). In each of the panels, the Gini importance of the top five
689 most discriminant biomarkers of microbial species, KOs, or metabolites are displayed.
690 Boxes represent the interquartile range between the first and third quartiles and the line
691 inside represents the median. Whiskers denote the lowest and highest values within the
692 1.5 × interquartile range from the first and third quartiles, respectively. Boxes are
693 marked in a specific color to show the significant elevation ($P < 0.05$, red, Mann-
694 Whitney-Wilcoxon test) or depletion ($P < 0.05$, blue, Mann-Whitney-Wilcoxon test) of
695 the biomarkers in each of the arthritis stages compared to the healthy group. The ROC
696 curves of the random-forest model using microbial species, KOs, or metabolites were
697 plotted, with AUC calculated by 10 randomized 10-fold cross-validation. The color of
698 the curve represents the category of the used features.

699

700 **Fig. 5 | Multi-omics dynamic patterns in the gut-joint axis across four RA stages.**
701 **a**, The alteration of gut microbiome taxonomic and functional composition would
702 interact with human joint through serum metabolites perturbation. The most dominant
703 elevation and depletion of microbial species, microbial function, and serum metabolites
704 in each of RA stages and OA are shown, as compared to those in the healthy group.
705 The perturbated metabolites and elevated serum cytokines such as IL-6 and TNF-α were
706 transferred to human joints to promote inflammation and bone erosion. Different RA
707 stages and OA had inflammation or bone erosion of different severity and
708 corresponding different microbial patterns. **b**, The dot plots show stage-specific
709 abundance or concentration (mean ± s.e.m.) of microbial species, serum cytokines, and
710 serum metabolites, which are specified in (a). Four RA stages are connected to display
711 the variance. For microbial species and serum metabolites, dots are colored differently
712 if the features are significantly elevated (red) or significantly depleted (blue), as
713 compared to those of the healthy group. For serum cytokines, dots of RA stages are
714 colored in red for significant elevation compared to the OA group. ***, $P < 0.005$; **,
715 $P < 0.01$; *, $P < 0.05$, Mann-Whitney-Wilcoxon test.

716

717 **Fig. 6 | Potential pathogenesis across successive RA stages from multi-omics**
718 **perspective. a**, Certain potential mechanisms by which gut microbial dysbiosis play
719 roles in RA pathogenesis through hurting hone tissue and increasing inflammation. The
720 driving species, microbial dysfunction, and related metabolites were extracted from **Fig.**
721 **3b**. The red or blue box of metabolites represents their speculated elevation or depletion
722 according to the KEGG map. The dotted line represents the speculated effects of
723 microbial and metabolic variation on arthritis pathogenesis, with a bracket including
724 the literature. **b**, The most representative effects of microbial dysbiosis and metabolic
725 disorders on RA progression across successive stages.

726 **Methods**

727 **The cohort for the metagenomic and metabolomic study**

728 A total of 122 fecal and 122 serum samples were collected from 122 outpatients of the
729 Shandong Provincial Qianfoshan Hospital (Jinan, Shandong, China). These outpatients
730 included 76 patients with rheumatoid arthritis (RA), 19 patients with osteoarthritis (OA),
731 and 27 healthy individuals. RA patients were grouped into four RA stages including
732 RAS1 (n = 15), RAS2 (n = 21), RAS3 (n = 18), and RAS4 (n = 22) according to the
733 rheumatoid diagnostic score[3], where RAS1, RAS2, RAS3, and RAS4 has a score of 6-
734 7, 8, 9, and 10, respectively (**Supplementary Tables 1**). The score was evaluated by
735 the sum of four categories[3]: A) joint involvement that refers to any swollen or tender

736 joint on examination; B) Serology determined by values of rheumatoid factor and anti-
737 citrullinated protein antibody; C) Acute-phase reactants determined by values of C-
738 reactive protein and erythrocyte sedimentation rate; and D) Duration of symptoms.
739 Subsequently, fecal samples were sequenced and serum samples were used to test the
740 serum metabolites and inflammatory cytokine levels. Serum metabolites were tested
741 using UHPLC-MS/MS, while inflammatory cytokines were quantified using the MESO
742 SCALE DISCOVERY (MSD) Quick Plex S600MM multiplex assay. Cytokine levels
743 of healthy individuals were not available because of their extremely low levels. All of
744 the participants were at fasting status during the sample collection in the morning. Only
745 participants who met the standard were recruited in this study: Recruited individuals
746 had not received treatment in the recent month, and had no malignant tumor, no other
747 rheumatic diseases such as ankylosing spondylitis, psoriasis, gout, no gastrointestinal
748 diseases such as diarrhea, constipation, and hematochezia in the recent one month, no
749 infections, no other comorbidity such as diabetes and hepatitis B.

750 **Metagenome sequencing and processing**

751 Whole-genome shotgun sequencing of fecal samples was carried out on the Illumina
752 Hiseq X Ten. All samples were paired-end sequenced with a 150-bp read length. Reads
753 containing more than 40 base pairs with low-quality values of 38 or less were discarded.
754 Reads containing more than 10 unidentified base pairs were discarded. Reads
755 containing more than 15 base pairs that overlapped adapter sequences were discarded.

756 Next, reads were mapped against the human genome using Bowtie2 (version 2.2.9).
757 Those that were mapped were considered to be derived from the human genome and
758 were discarded. Finally, unpaired reads were discarded. High-quality reads were used
759 for the following analyses. MetaPhlAn2[19] was then used to generate microbial species
760 profiles.

761 **Contig assembly and ORF prediction**

762 The reads were assembled into contigs using MEGAHIT[18] (version 1.2.6) with the
763 minimum contig length set at 500 bp and default parameters for metagenomics. The
764 open reading frames (ORFs) were predicted from the assembled contigs using
765 Prodigal[51] (version 2.6.3) with default parameters. The ORFs of <100 bp were removed.

766 **Non-redundant gene catalogue construction**

767 The ORFs were clustered to remove redundancy using Cd-hit[52] (version 4.6.6) with a
768 sequence identity threshold set at 0.95 and the alignment coverage set at 0.9, which
769 resulted in a catalogue of 4,047,645 nonredundant genes. The non-redundant genes
770 were then collapsed into metagenomic species (MGS)[17,53], and grouped into KEGG
771 functional modules[17], as described below.

772 **Identification of metagenomic species**

773 High-quality reads were mapped to the catalogue of nonredundant genes using Bowtie
774 2[54] (version 2.2.9) with default parameters. The abundance profile for each catalogue

775 gene was calculated as the sum of uniquely mapped sequence reads, using 19M
776 sequence reads per sample (downsized). The co-abundance clustering of the 4,047,645
777 genes was performed using canopy algorithm[53] (http://git.dworzynski.eu/mgs-canopy-
778 algorithm), and 553 gene clusters that met the previously described criteria[53] and
779 contained more than 700 genes were referred to as MGS. MGS that were present in at
780 least 4 samples were used for the following analysis. The abundance profiles of MGS
781 were determined as the medium gene abundance throughout the samples. MGS were
782 taxonomically annotated by summing up the taxonomical annotation of their genes as
783 described by Nielsen *et al*[53]. Each MGS gene was annotated by sequence similarity to
784 the bacterial complete genome in NCBI Reference Sequence Database (BLASTN, E-
785 value < 0.001).

786 **Annotation of KEGG modules.**

787 The catalogue of the nonredundant genes was functionally annotated to the KEGG
788 database (release 94.0) by KofamKOALA (version 1.3.0)[55,56]. The produced KEGG
789 Orthologs (KOs) were mapped to the KEGG modules annotation downloaded on 1
790 August 2020 from the KEGG BRITE database. KOs that were present in at least 4
791 samples were used for the following analysis. The KO abundance profile was calculated
792 by summing the abundance of genes that were annotated to each of the KOs.

793 **Leave-one-out analysis**

794 Leave-one-out analysis was used to test which MGS was driving the observed
795 associations between KEGG modules and arthritis. The calculation of the KO
796 abundance was iterated excluding the genes from a different MGS, in each iteration.
797 The effect of a given MGS on a specified association was defined as the change in
798 median Spearman correlation coefficient between KOs and arthritis when genes from
799 the respective MGS were left out, as previously described[17,57].

800 **Multi-omics biomarkers**

801 A random forest algorithm was performed using the R package "randomForest" on each
802 of the three profiles independently, including 232 microbiome species, 6,224 KOs, and
803 277 metabolites. Function "trainControl" in R package "caret" was used to perform 10
804 repeats of 10-fold cross-validation for each data set. Function "train" in R package
805 "caret" was used to fid models over different tuning parameters to determine the "mtry"
806 for random forest algorithm.

807 **Correlation of phenotypes and KEGG modules**

808 The clinical phenotypes including types of arthritis (Healthy = 0, OA = 1, RA = 2) were
809 used in the correlation analysis with KEGG modules. The associations of clinical
810 phenotypes with KEGG modules were determined by if Spearman correlations of the
811 phenotype with the abundances of KOs in the given KEGG module were significantly

812 higher or lower (Mann–Whitney U-test FDR < 0.1) than with the abundances of all
813 other KOs. The phenotypes adjusted by age and gender were also tested using partial
814 Spearman correlation. The union set of the significant associations between KEGG
815 modules and phenotypes/phenotypes adjusted by age and gender were used for the
816 following analysis.

817 **Gavage experiments using ascorbate in the CIA mouse model**

818 Nine healthy seven-week-old DBA/1 mice weighing 20g were fed in an ultra-clean
819 animal laboratory (SPF grade) with a humidity of 55% and a temperature of 26°C. CIA
820 models were constructed and established as described before[58]. Nine mice were then
821 divided into three groups (three mice per group), including normal DBA/1 mice and
822 two groups of DBA/1 mice with CIA. Three-month gavage (0.3ml/d) to three groups
823 of mice was conducted, including 1) 0.9% normal saline to normal DBA/1 mice, 2) 0.9%
824 normal saline to DBA/1 mice with CIA, and 3) 100ng/ul ascorbate to DBA/1 mice with
825 CIA. After three-month gavage, the mice serum TNF-α level and the IL-6 level were
826 tested using ELISA kit (mlbio, China). Mice were then killed and preserved in
827 4% formalin for two days. Micro-CT (QuantumGX, PerkinElmer, UnitedStates) was
828 used to perform scanning and three-dimensional structural reconstruction of the joints.
829 The settings were set to 209m, 90kV X-ray tube voltage, 160uA of the current, and 3
830 minutes of the scan time. The angle of the X-ray scan rotated 180 degrees.

831 **The cohort for detection of bacteria in the joint synovial fluid**

832 A total of 271 patients with RA of four distinct stages were recruited, including 52
833 RAS1, 66 RAS2, 67 RAS3, and 86 RAS4. RA patients were grouped into four RA
834 stages according to the rheumatoid diagnostic score[3]. The patients met the
835 aforementioned standard. Synovial- fluid (SF) samples were collected aseptically from
836 knee joints during therapeutic aspiration. Synovial-fluidSF samples were deposited in
837 sterile tubes on ice and homogenized within five minutes of collection. A tube filled
838 with sterile phosphate-buffered saline (PBS) was left open throughout the procedure
839 and subsequently processed in parallel with the samples as a negative control. The
840 entire experiment was conducted in a completely sterile atmosphere. Each sample was
841 immediately frozen and kept without heparin or hyaluronidase at –135 °C. For each
842 patient, a total of 7ml synovial fluidSF was collected, of which 5ml was utilized for
843 16S rRNA gene sequencing, 1ml was used for bacteria isolation, and 1ml synovial fluid
844 SF was prepared for scanning electron microscopy.

845 **16S rRNA gene sequencing and processing**

846 Bacterial DNA was extracted from 271 5ml synovial fluid samples. The tube containing
847 PBS serves as environmental control. 86 synovial-fluidSF samples from patients in
848 RAS4 had bacteria DNA content (≥10ng) (Bacterial DNA Kit, TIANGEN) for bacteria
849 16S rRNA gene high-throughput sequencing. The V1/V2 hypervariable regions of the
850 16S ribosomal RNA gene were sequenced using the Illumina HiSeq platform. The 16S

851 sequence paired-end data set was joined and quality filtered using the FLASH as
852 previously described[59]. Sequence analysis was performed using the Quantitative
853 Insights into Microbial Ecology (QIIME, version 1.9.1) software[60]. Chimeric sequences
854 were removed using de novo chimera detection of USEARCH 6.1[61]. Sequences were
855 clustered against the 2013 Greengenes (13_8 release) ribosomal database's 97%
856 reference data set. Sequences that did not match any entries in this reference were
857 subsequently clustered into de novo OTUs at 97% similarity with UCLUST[61].
858 Taxonomy was assigned to all OTUs using the RDP classifier[62] against Greengenes
859 reference data set.

860 **Isolation of bacteria in the joint synovial fluid**

861 Due to that pH of joint synovial fluidSF was 7.6, we used Luria-Bertani (LB) broth
862 medium with a pH of 7.6 for cultivation. The procedures were as follows: 1) sterilized
863 LB medium (pH7.6) and inoculation equipment were placed in an anaerobic glove box
864 (Ruskinn Concept 400) for deaeration one day in advance; 2) 1ml synovial-fluid SF
865 samples per stage of RA were used for bacterial cultivation. 3) The synovial-fluid SF
866 sample was serially diluted with sterilized water, plated onto LB (0.5% yeast extract,
867 1% tryptone, 1% sodium chloride) agar medium, and then incubated at 37°C for 72 h
868 to obtain single colonies. Then, three colonies per plate were randomly selected and
869 streaked three consecutive times on LB agar medium to obtain a pure culture, which
870 was named isolate SF1 to SF9, respectively. To identify isolate SF1 to SF9, a partial

871 fragment of 16S rDNA was amplified with the primer pair 27F (5'-
872 AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-
873 3'), and then DNA sequencing was performed for preliminary identification.

874 **Scanning electron microscopy for bacteria in the joint synovial fluid**

875 The synovial-fluidSF sample was filtered through the membrane (special membrane for
876 flow cytometry) to remove solids and large particles (Filtrate A). Filtrate A was then
877 filtered through a 0.45 μmum membrane to remove most human cells (Filtrate B).
878 Filtrate B was then filtered through a 0.22 μmum filter membrane to enrich bacteria and
879 subsequently washed with sterilized ultra-pure water to obtain Filtrate C. Filtrate C was
880 soaked in 2.5% glutaraldehyde for 4 hours at room temperature. Filtrate C was then
881 washed three times with 0.1 M PBS buffer and treated with 1% osmic acid for 4h.
882 Filtrate C was then dehydrated in ethanol, vacuum dried by tert-butyl alcohol, coated
883 with gold, and imaged with a scanning electron microscope (ZEISS Sigma 300).

884 **Statistical analyses**

885 The correlations between vectors such as microbial species, metabolites, and the levels
886 of cytokines, were tested using Spearman correlation or partial Spearman correlation
887 with age and gender adjusted (*P* values), corrected by FDR (q values). For categorical
888 metadata, samples were pooled into bins (RA/OA/Healthy, RAS1/RAS2/RAS3/RAS4)
889 and significant features were identified using Mann-Whitney-Wilcoxon Test (*P* values)
890 with Benjamini and Hochberg correction (FDR, q values).

891 **Data availability**

892 Whole-genome shotgun sequencing data are available in the Genome Sequence
893 Archive (GSA) section of the National Genomics Data Center (project accession
894 number CRA004348). 16S rRNA gene sequencing data are available in the Genome
895 Sequence Archive (GSA) section of the National Genomics Data Center (project
896 accession number CRA005811).

897 **Methods references**

898 51. Hyatt, D., *et al.* Prodigal: prokaryotic gene recognition and translation initiation
899 site identification. *BMC Bioinform* **11**, 119 (2010).
900 52. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large
901 sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
902 53. Nielsen, H.B., *et al.* Identification and assembly of genomes and genetic
903 elements in complex metagenomic samples without using reference genomes.
904 *Nat. Biotechnol.* **32**, 822–828 (2014).
905 54. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat.*
906 *Methods* **9**, 357–359 (2012).
907 55. Kanehisa, M., *et al.* Data, information, knowledge and principle: back to
908 metabolism in KEGG. *Nucleic Acids Res.* **42**, D199–205 (2014).
909 56. Aramaki, T., *et al.* KofamKOALA: KEGG Ortholog assignment based on
910 profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252

(2020).

57. Pedersen, H.K., *et al.* Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **535**, 376–381 (2016).

58. Song, G., *et al.* CD109 regulates the inflammatory response and is required for the pathogenesis of rheumatoid arthritis. *Ann. Rheum. Dis.* **78**, 1632–1641 (2019).

59. Magoč, T. & Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

60. Caporaso, J.G., *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).

61. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

62. Cole, J.R., *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37**, D141–145 (2009).

925 **Acknowledgements**

926 This work was partially supported by the National Natural Science Foundation of China
927 (grant numbers 31871334, 82003766, 32071465, and 31671374), the Academic
928 Promotion Project of Shandong First Medical University (grant number 2019LJ001),
929 and the Key Research and Development project of Shandong Province (grant number
930 2021ZDSYS27).

931 **Author contributions**

932 M.C., Y.Z., L.Z., K.N., and J.H. designed the study. M.C., Y.Z., Y.C., Y.Zha, S.L.,
933 G.C., and M.L. collected samples and conducted experiments. M.C., Y.Z, and K.N.
934 conducted data analysis and produced the figures and tables. M.C., Y.Z., K.N., and J.H.
935 wrote the manuscript. All authors revised the manuscript. M.C., Y.Z., L.Z., K.N., and
936 J.H. supervised the study.

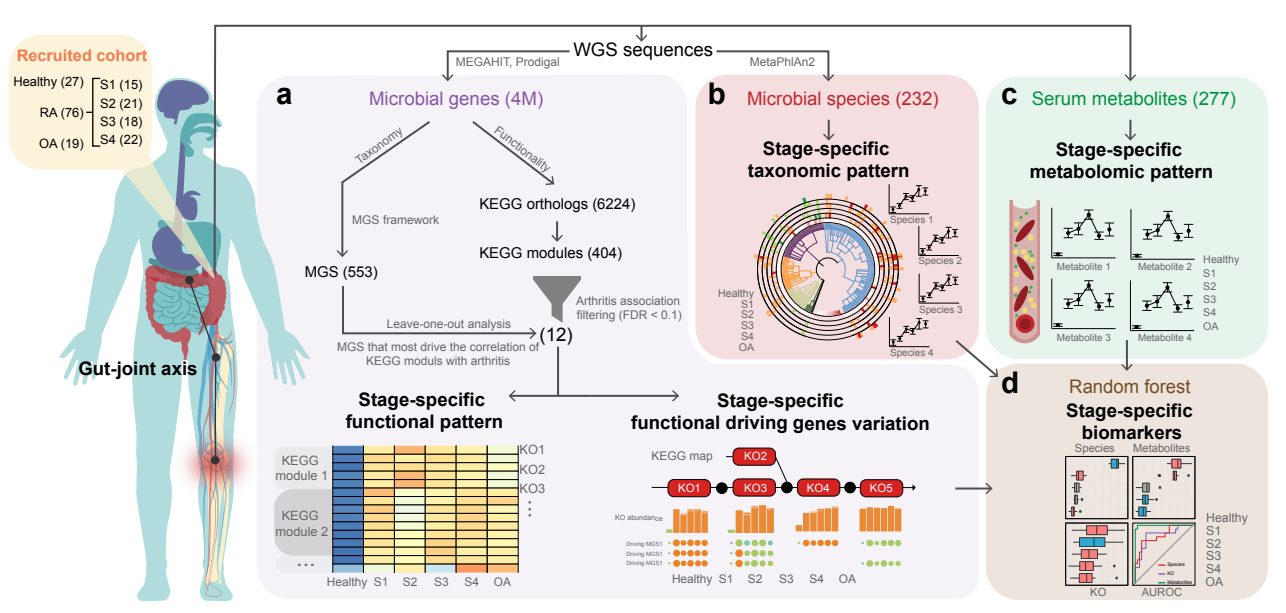
937 **Ethics statement**

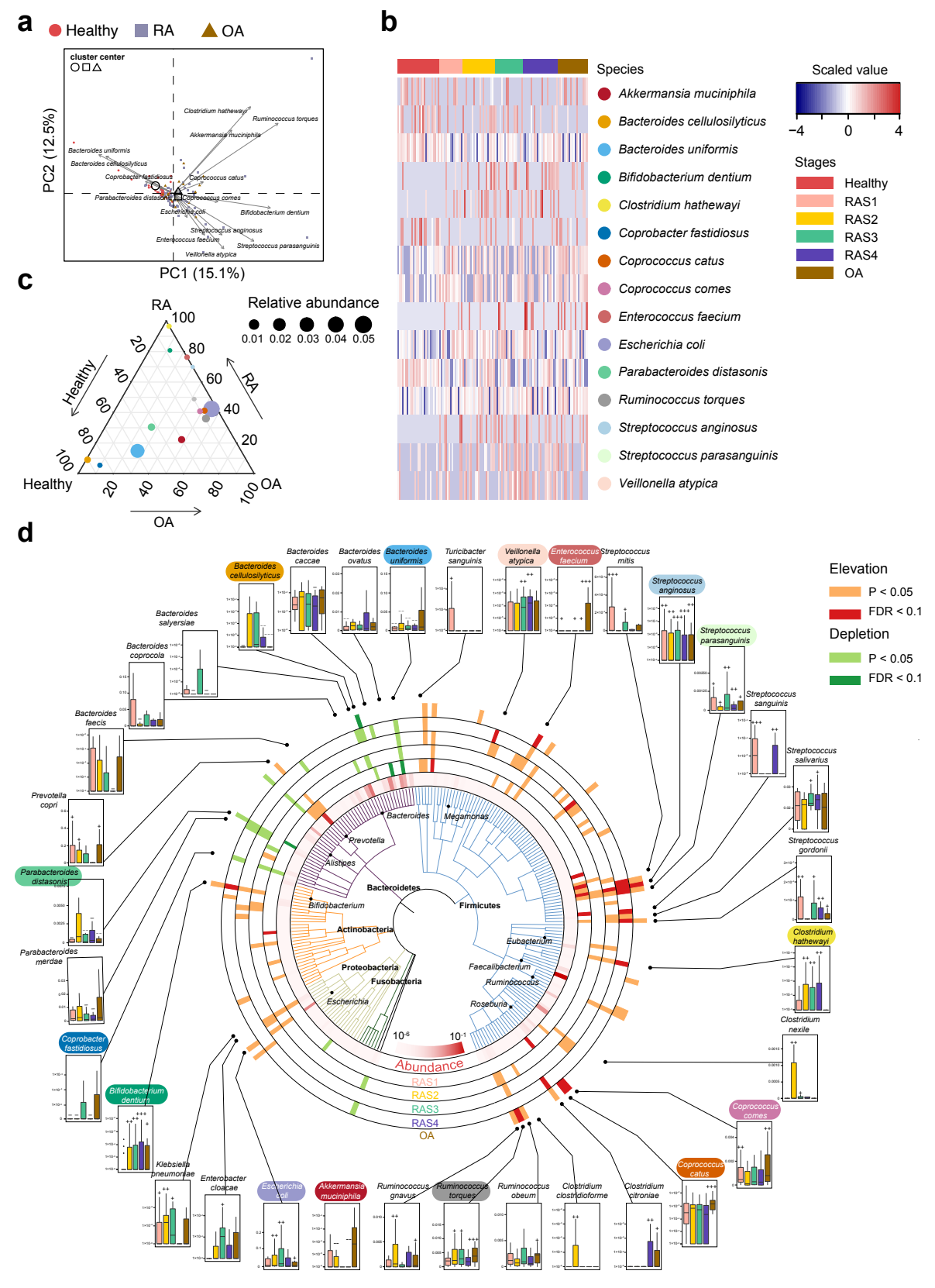
938 The study was approved by the Ethics Committee of The First Affiliated Hospital of
939 Shandong First Medical University (NO.2017-02).

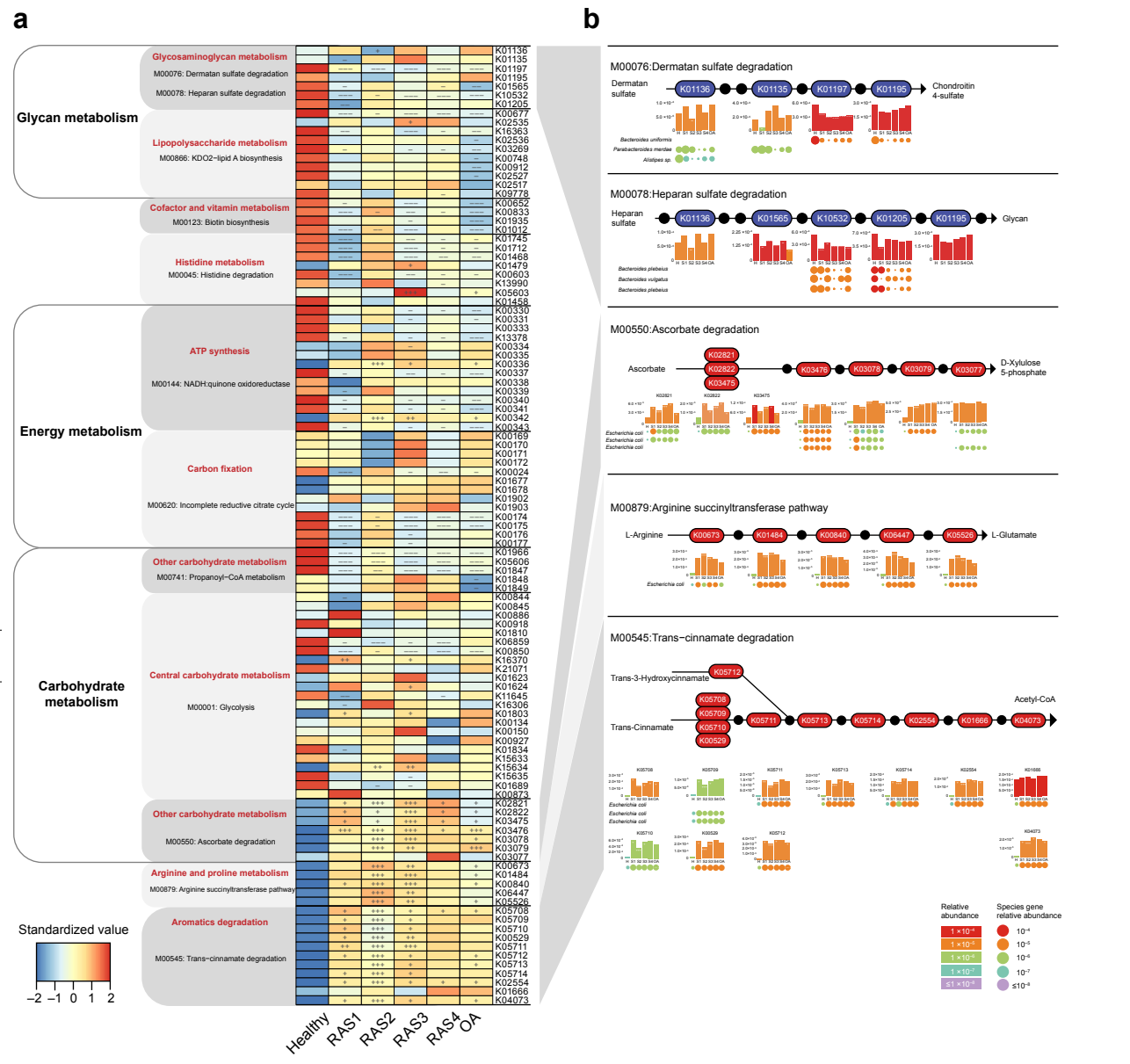
940 **Competing interests**

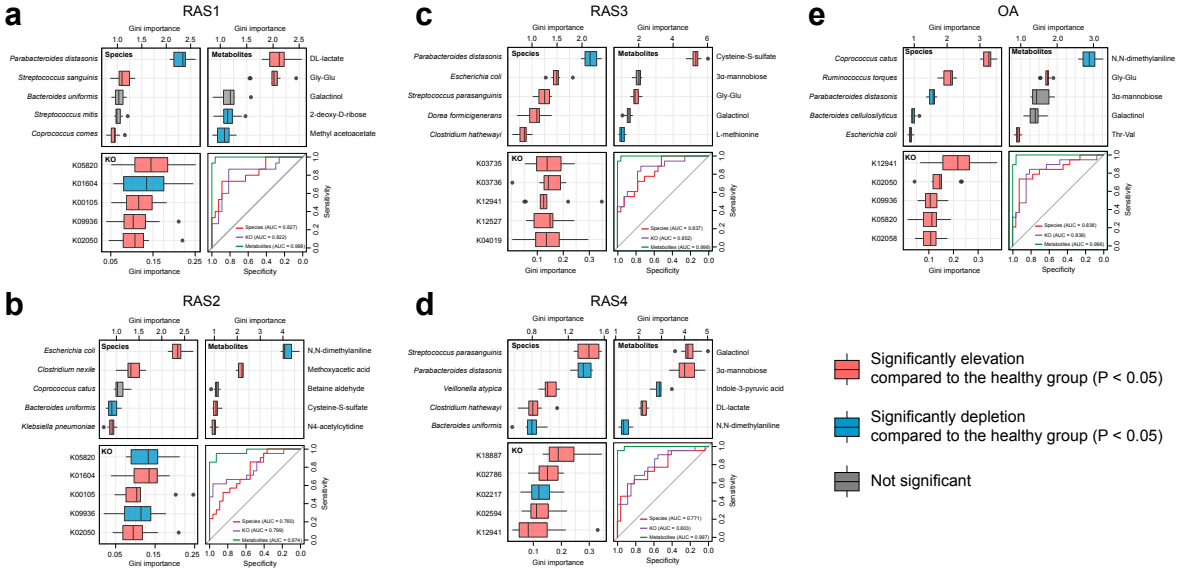
941 The authors declare that they have no competing interests

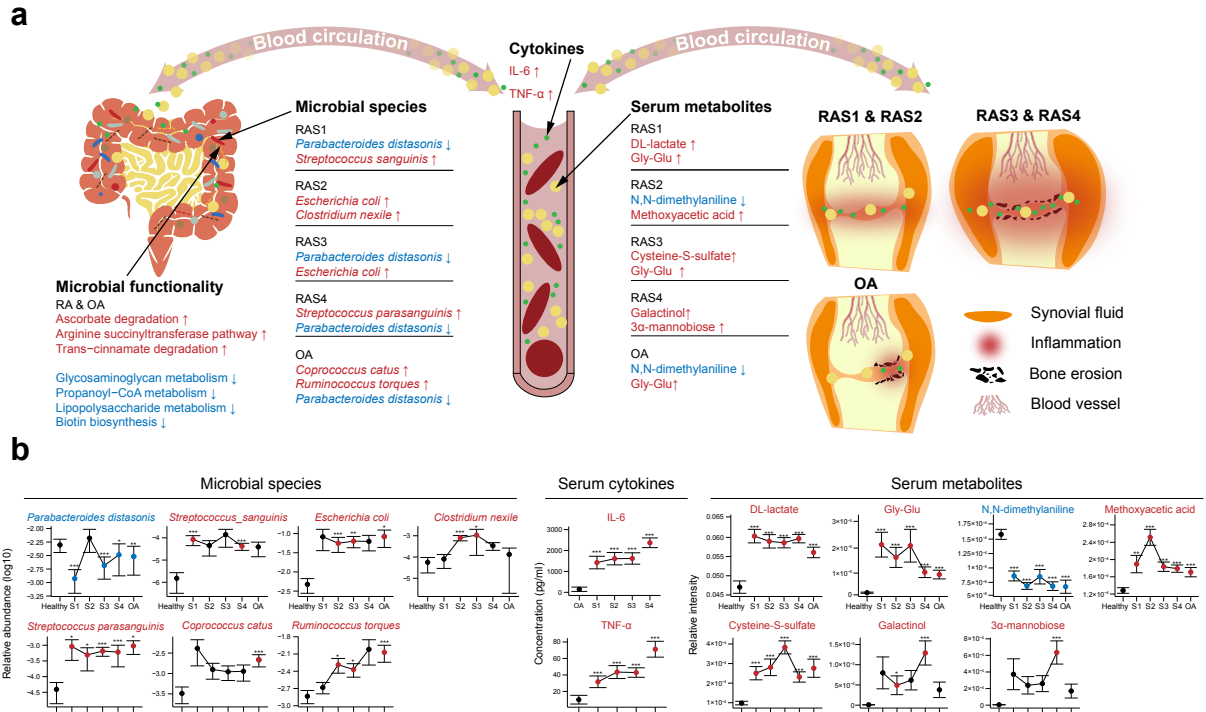
942

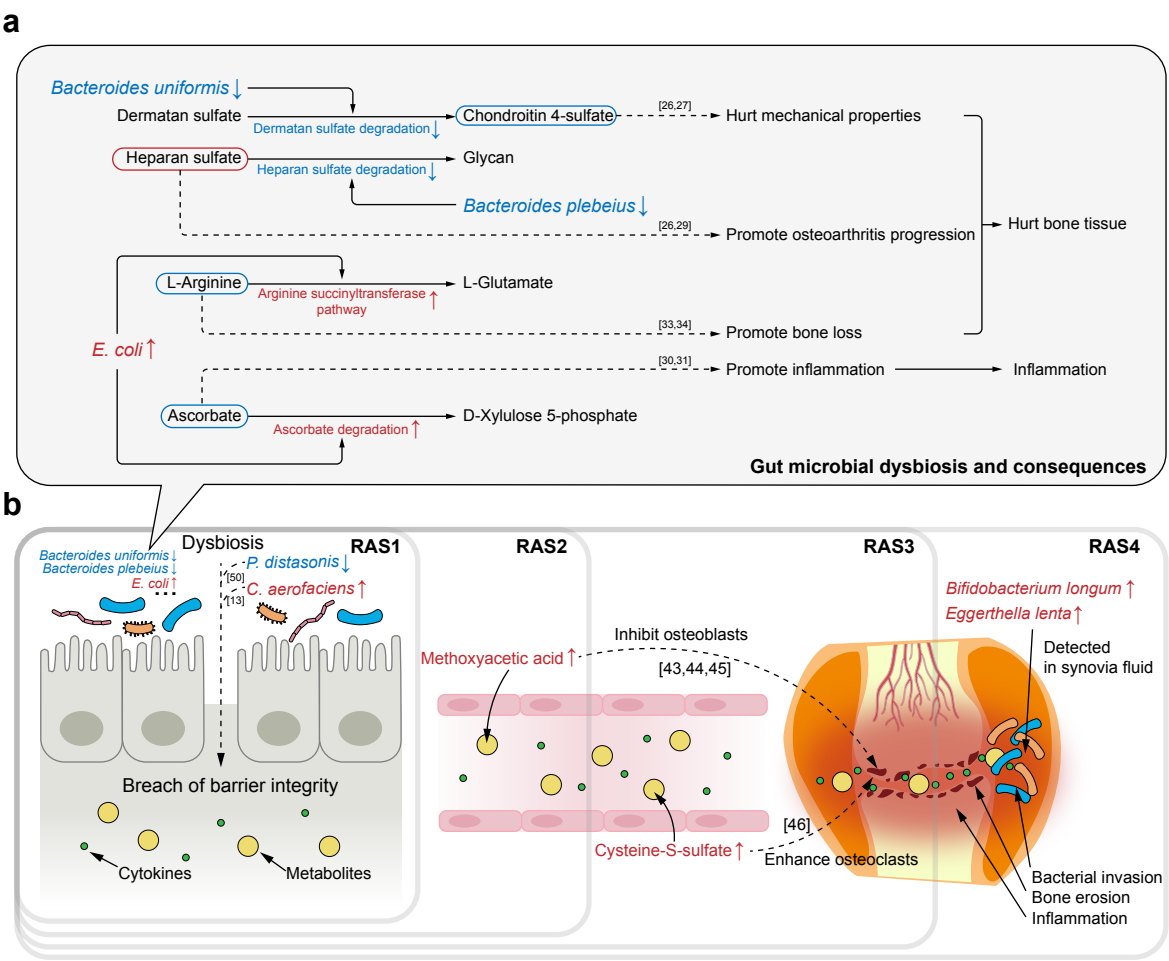

Recruited cohort
Healthy (27)
RA (76)
S1 (15)
S2 (21)
S3 (18)
S4 (22)
OA (19)
Gut-joint axis
WGS sequences
MEGAHIT, Prodigal
MetaPhlAn2
a
Microbial genes (4M)
Taxonomy
Functionality
MGS framework
MGS (553)
KEGG orthologs (6224)
KEGG modules (404)
Arthritis association filtering (FDR < 0.1)
Leave-one-out analysis
MGS that most drive the correlation of KEGG moduls with arthritis
(12)
Stage-specific functional pattern
KEGG module 1
KEGG module 2
KO1
KO2
KO3
Healthy S1 S2 S3 S4 OA
Stage-specific functional driving genes variation
KEGG map
KO1
KO2
KO3
KO4
KO5
KO abundance
Driving MGS1
Healthy S1 S2 S3 S4 OA
b
Microbial species (232)
Stage-specific taxonomic pattern
Species 1
Species 2
Species 3
Species 4
Healthy
S1
S2
S3
S4
OA
c
Serum metabolites (277)
Stage-specific metabolomic pattern
Metabolite 1
Metabolite 2
Metabolite 3
Metabolite 4
Healthy
S1
S2
S3
S4
OA
d
Random forest
Stage-specific biomarkers
Species
Metabolites
KO
AUROC
Species
KO
Metabolites
Healthy
S1
S2
S3
S4
OA


a
Healthy
RA
OA
cluster center
PC2 (12.5%)
PC1 (15.1%)
Clostridium hathewayi
Ruminococcus torques
Akkermansia muciniphila
Bacteroides uniformis
Bacteroides cellulosilyticus
Coprobacter fastidiosus
Coprococcus catus*
Parabacteroides distasonis
Coprococcus comes
Escherichia coli
Bifidobacterium dentium
Streptococcus anginosus
Enterococcus faecium
Streptococcus parasanguinis
Veillonella atypica
b
Species
Akkermansia muciniphila
Bacteroides cellulosilyticus
Bacteroides uniformis
Bifidobacterium dentium
Clostridium hathewayi
Coprobacter fastidiosus
Coprococcus catus
Coprococcus comes
Enterococcus faecium
Escherichia coli
Parabacteroides distasonis
Ruminococcus torques
Streptococcus anginosus
Streptococcus parasanguinis
Veillonella atypica
Scaled value
−4
0
4
Stages
Healthy
RAS1
RAS2
RAS3
RAS4
OA
c
RA
Healthy
OA
Relative abundance
0.01
0.02
0.03
0.04
0.05
d
Elevation
P < 0.05
FDR < 0.1
Depletion
P < 0.05
FDR < 0.1
Bacteroides cellulosilyticus
Bacteroides caccae
Bacteroides ovatus
Bacteroides uniformis
Turicibacter sanguinis
Veillonella atypica
Enterococcus faecium
Streptococcus mitis
Streptococcus anginosus
Streptococcus parasanguinis
Streptococcus sanguinis
Streptococcus salivarius
Streptococcus gordonii
Clostridium hathewayi
Clostridium nexile
Coprococcus comes
Coprococcus catus
Clostridium citroniae
Clostridium clostridioforme
Ruminococcus obeum
Ruminococcus torques
Ruminococcus gnavus
Akkermansia muciniphila
Escherichia coli
Enterobacter cloacae
Klebsiella pneumoniae
Bifidobacterium dentium
Coprobacter fastidiosus
Parabacteroides merdae
Parabacteroides distasonis
Prevotella copri
Bacteroides faecis
Bacteroides coprocola
Bacteroides salyersiae
Bacteroides
Megamonas
Prevotella
Alistipes
Bacteroidetes
Firmicutes
Bifidobacterium
Actinobacteria
Eubacterium
Proteobacteria
Faecalibacterium
Fusobacteria
Ruminococcus
Escherichia
Roseburia
10^{-6}
10^{-1}
Abundance
RAS1
RAS2
RAS3
RAS4
OA


a
Glycan metabolism
Glycosaminoglycan metabolism
M00076: Dermatan sulfate degradation
M00078: Heparan sulfate degradation
Lipopolysaccharide metabolism
M00866: KDO2−lipid A biosynthesis
Cofactor and vitamin metabolism
M00123: Biotin biosynthesis
Histidine metabolism
M00045: Histidine degradation
Energy metabolism
ATP synthesis
M00144: NADH;quinone oxidoreductase
Carbon fixation
M00620: Incomplete reductive citrate cycle
Carbohydrate metabolism
Other carbohydrate metabolism
M00741: Propanoyl−CoA metabolism
Central carbohydrate metabolism
M00001: Glycolysis
Other carbohydrate metabolism
M00550: Ascorbate degradation
Arginine and proline metabolism
M00879: Arginine succinyltransferase pathway
Aromatics degradation
M00545: Trans−cinnamate degradation
K01136
K01135
K01197
K01195
K00677
K02535
K16363
K02536
K03269
K00748
K00912
K02527
K02517
K09778
K00652
K00833
K01935
K01012
K01745
K01712
K01468
K01479
K00603
K13990
K05603
K01458
K00330
K00331
K00333
K13378
K00334
K00335
K00336
K00337
K00338
K00339
K00340
K00341
K00342
K00343
K00169
K00170
K00171
K00172
K00024
K01677
K01678
K01902
K01903
K00174
K00175
K00176
K00177
K01966
K05606
K01847
K01848
K01849
K00844
K00845
K00886
K00918
K01810
K06859
K06850
K16370
K21071
K01623
K01624
K11645
K16306
K01803
K00134
K00150
K00927
K01834
K15633
K15634
K15635
K01689
K00873
K02821
K02822
K03475
K03476
K03078
K03079
K03077
K00673
K01484
K00840
K06447
K05526
K05708
K05709
K05710
K00529
K05711
K05712
K05713
K05714
K02554
K01666
K04073
Standardized value
−2 −1 0 1 2
Healthy
RAS1
RAS2
RAS3
RAS4
OA
b
M00076:Dermatan sulfate degradation
Dermatan sulfate
Chondroitin 4-sulfate
Bacteroides uniformis
Parabacteroides merdae
Alistipes sp.
M00078:Heparan sulfate degradation
Heparan sulfate
Glycan
Bacteroides plebeius
Bacteroides vulgatus
Bacteroides plebeius
M00550:Ascorbate degradation
Ascorbate
D-Xylulose 5-phosphate
Escherichia coli
Escherichia coli
Escherichia coli
M00879:Arginine succinyltransferase pathway
L-Arginine
L-Glutamate
Escherichia coli
M00545:Trans−cinnamate degradation
Trans-3-Hydroxycinnamate
Trans-Cinnamate
Acetyl-CoA
Escherichia coli
Escherichia coli
Escherichia coli
Relative abundance
Species gene relative abundance


a
RAS1
Gini importance
Species
Parabacteroides distasonis
Streptococcus sanguinis
Bacteroides uniformis
Streptococcus mitis
Coprococcus comes
Metabolites
DL-lactate
Gly-Glu
Galactinol
2-deoxy-D-ribose
Methyl acetoacetate
KO
K05820
K01604
K00105
K09936
K02050
Species (AUC = 0.827)
KO (AUC = 0.820)
Metabolites (AUC = 0.968)
Sensitivity
Specificity
b
RAS2
Escherichia coli
Clostridium nexile
Coprococcus catus
Bacteroides uniformis
Klebsiella pneumoniae
N,N-dimethylaniline
Methoxyacetic acid
Betaine aldehyde
Cysteine-S-sulfate
N4-acetylcytidine
Species (AUC = 0.760)
KO (AUC = 0.799)
Metabolites (AUC = 0.974)
c
RAS3
Parabacteroides distasonis
Escherichia coli
Streptococcus parasanguinis
Dorea formicigenerans
Clostridium hathewayi
Cysteine-S-sulfate
3α-mannobiose
Gly-Glu
Galactinol
L-methionine
K03735
K03736
K12941
K12527
K04019
Species (AUC = 0.837)
KO (AUC = 0.832)
Metabolites (AUC = 0.966)
d
RAS4
Streptococcus parasanguinis
Parabacteroides distasonis
Veillonella atypica
Clostridium hathewayi
Bacteroides uniformis
Galactinol
3α-mannobiose
Indole-3-pyruvic acid
DL-lactate
N,N-dimethylaniline
K18887
K02786
K02217
K02594
K12941
Species (AUC = 0.771)
KO (AUC = 0.803)
Metabolites (AUC = 0.987)
e
OA
Coprococcus catus
Ruminococcus torques
Parabacteroides distasonis
Bacteroides cellulosilyticus
Escherichia coli
N,N-dimethylaniline
Gly-Glu
3α-mannobiose
Galactinol
Thr-Val
K12941
K02050
K09936
K05820
K02058
Species (AUC = 0.838)
KO (AUC = 0.838)
Metabolites (AUC = 0.995)
Significantly elevation compared to the healthy group (P < 0.05)
Significantly depletion compared to the healthy group (P < 0.05)
Not significant


a
Blood circulation
Blood circulation
Cytokines
IL-6 ↑
TNF-α ↑
Microbial species
RAS1
Parabacteroides distasonis ↓
Streptococcus sanguinis ↑
RAS2
Escherichia coli ↑
Clostridium nexile ↑
RAS3
Parabacteroides distasonis ↓
Escherichia coli ↑
RAS4
Streptococcus parasanguinis ↑
Parabacteroides distasonis ↓
OA
Coprococcus catus ↑
Ruminococcus torques ↑
Parabacteroides distasonis ↓
Serum metabolites
RAS1
DL-lactate ↑
Gly-Glu ↑
RAS2
N,N-dimethylaniline ↓
Methoxyacetic acid ↑
RAS3
Cysteine-S-sulfate↑
Gly-Glu ↑
RAS4
Galactinol↑
3α-mannobiose ↑
OA
N,N-dimethylaniline ↓
Gly-Glu↑
RAS1 & RAS2
RAS3 & RAS4
OA
Microbial functionality
RA & OA
Ascorbate degradation ↑
Arginine succinyltransferase pathway ↑
Trans−cinnamate degradation ↑
Glycosaminoglycan metabolism ↓
Propanoyl−CoA metabolism ↓
Lipopolysaccharide metabolism ↓
Biotin biosynthesis ↓
Synovial fluid
Inflammation
Bone erosion
Blood vessel
b
Microbial species
Serum cytokines
Serum metabolites
Parabacteroides distasonis
Streptococcus_sanguinis
Escherichia coli
Clostridium nexile
Streptococcus parasanguinis
Coprococcus catus
Ruminococcus torques
Relative abundance (log10)
Healthy S1 S2 S3 S4 OA
IL-6
TNF-α
Concentration (pg/ml)
DL-lactate
Gly-Glu
N,N-dimethylaniline
Methoxyacetic acid
Cysteine-S-sulfate
Galactinol
3α-mannobiose
Relative intensity


a
Bacteroides uniformis ↓
Dermatan sulfate
Dermatan sulfate degradation ↓
Chondroitin 4-sulfate
[26,27]
Hurt mechanical properties
Heparan sulfate
Heparan sulfate degradation ↓
Glycan
Bacteroides plebeius ↓
[26,29]
Promote osteoarthritis progression
Hurt bone tissue
L-Arginine
Arginine succinyltransferase pathway ↑
L-Glutamate
[33,34]
Promote bone loss
E. coli ↑
[30,31]
Promote inflammation
Inflammation
Ascorbate
Ascorbate degradation ↑
D-Xylulose 5-phosphate
Gut microbial dysbiosis and consequences
b
Dysbiosis
RAS1
Bacteroides uniformis ↓
Bacteroides plebeius ↓
E. coli ↑
P. distasonis ↓
[50]
[13]
C. aerofaciens ↑
Breach of barrier integrity
Cytokines
Metabolites
RAS2
RAS3
RAS4
Methoxyacetic acid ↑
Inhibit osteoblasts
[43,44,45]
Cysteine-S-sulfate ↑
[46]
Enhance osteoclasts
Bifidobacterium longum ↑
Eggerthella lenta ↑
Detected in synovia fluid
Bacterial invasion
Bone erosion
Inflammation