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

3 Mingyue Cheng^{1,2†}, Yan Zhao^{1†}, Yazhou Cui¹, Yuguo Zha², Shufeng Li^{1,3},
4 Guangxiang Cao¹, Mian Li¹, Jinxiang Han^{1*}, Kang Ning^{2*}

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

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

12 ³Shandong Key Laboratory of Rheumatic Disease and Translational Medicine, Department of
13 Orthopedic Surgery, The First Affiliated Hospital of Shandong First Medical University &
14 Shandong Provincial Qianfoshan Hospital, Jinan, China

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

16 ***Correspondence:**

17 Jinxiang Han (jxhan@sdfmu.edu.cn)

18 Kang Ning (ningkang@hust.edu.cn)

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

21

22 Abstract

23 Rheumatoid arthritis (RA) is a chronic inflammatory disease that typically progresses
 24 through four stages, from mild stiffness to severe disability. The gut-joint axis is
 25 critical in RA pathogenesis, while its dynamic patterns and regulatory mechanisms
 26 throughout RA progression remain unknown. Here, we performed fecal metagenomic
 27 and serum metabolomic studies on a cohort of 122 patients from four distinct RA
 28 stages. Microbial dysbiosis and metabolomic disorders present stage-specific patterns
 29 and played crucial roles in RA pathogenesis across successive stages. In particular,
 30 *Escherichia coli* drove increased ascorbate degradation particularly in the second and
 31 third stages, promoting inflammation and bone destruction. Besides, abnormally high
 32 levels of serum methoxyacetic acid and cysteine-S-sulfate inhibited osteoblasts in the
 33 second stage and enhanced osteoclasts in the third stage, respectively. Furthermore,
 34 combined with analysis of 16S rRNA gene sequencing and scanning electron
 35 microscopy on joint synovial fluid from another cohort of 271 RA patients, microbial
 36 invasion of the joint was very likely to happen in the fourth stage. Our findings
 37 elucidate the role of microbial dysbiosis and metabolic disorders in successive stages
 38 of RA, which open up new avenues for RA prognosis and therapy.

39 Introduction

40 Rheumatoid arthritis (RA) affects over tens of millions of people worldwide¹. RA is
 41 recognized clinically as a chronic, inflammatory, autoimmune disease that has four
 42 stages in progression^{2,3}. In the first stage, the synovium of the joint is inflamed. Most
 43 people have minor symptoms such as stiffness upon awakening². In the second stage,
 44 the inflamed synovium has caused damage to the joint cartilage. People begin to feel
 45 swelling, and have a restricted range of motion². In the third stage, RA has proceeded
 46 to a severe state when bone erosion begins. By this stage, the cartilage on the surface
 47 of the bones has deteriorated, resulting in the bones rubbing against one another.
 48 People feel increased swelling and mobility problems². In the fourth stage, certain
 49 joints were severely deformed and lost function. People suffer considerably from

50 swelling, stiffness, mobility loss, and disability². To inhibit RA progression, specific
51 therapeutic strategies are necessary for people across different RA stages.

52 Gut microbial dysbiosis has been implicated in the pathogenesis of RA, designated as
53 gut-joint axis, via a range of mechanisms such as metabolic perturbation and immune
54 response regulation⁴⁻⁹. For instance, increased abundance of *Prevotella* and
55 *Collinsella* in RA patients are correlated with the production of T_H17 cell
56 cytokines^{10,11}. Additionally, mice models have confirmed microbial roles in RA.
57 Reduction of *Bacteroides* spp. in mice with collagen-induced arthritis (CIA) may
58 contribute to a local pro-inflammatory environment by inhibiting CD4+ T cell
59 differentiation into regulatory T cells¹². Microbial metabolites have also been
60 correlated to immunity regulation in RA. Short-chain fatty acids (SCFAs) have a local
61 protective effect on colonic inflammation, via binding to FFAR2 on colonic T cells
62 and ILC3s^{8,13}. SCFA administration to mice with CIA can reduce the severity of
63 arthritis by modulation of IL-10^{5,8}. Comprehensive metagenomic and metabolomic
64 analyses may therefore enhance our understanding about the gut-joint axis. The role
65 of the gut-joint axis across successive stages of RA is understudied^{5,14}, where more
66 examination may provide an alternative approach to ameliorate RA progression.

67 In this study, we performed fecal metagenomic and serum metabolomic studies on
68 four stages of RA to depict the stage-specific patterns of microbial dysbiosis and
69 metabolic disorders. Our study demonstrates that microbial dysbiosis and metabolic
70 disorders plays continuous roles in RA pathogenesis across successive stages. In
71 particular, the depletion of *Bacteroides uniformis* and *Bacteroides plebeius* drove
72 microbial dysfunction in glycosaminoglycan metabolism, continuously hurting
73 articular cartilage. The elevation of *Escherichi coli* drove microbial dysfunction
74 mostly in the second and third stages, promoting inflammation and bone lose.
75 Moreover, metabolome was extremely disordered in the whole progression and
76 induced bone erosion manly through the abnormal elevation of methoxyacetic acid in
77 the second stage and cysteine-S-sulfate in the third stage. Furthermore, together with
78 analysis of 16S rRNA gene sequencing and scanning electron microscopy on joint

79 synovial fluid from another cohort of 271 RA patients, we concluded that the buildup
80 of these effects might lead to microbial invasion of the joint in the fourth stage of RA:
81 Gut microbes *Eggerthella lenta* and *Bifidobacterium longum* were elevated
82 exclusively in the fourth stage, and they could also be detected in the joint synovial
83 fluid of patients in the fourth stage. Collectively, these findings may open new
84 opportunities for prognosis, pain relief, and therapies throughout RA progression.

85 **Results**

86 **Multi-omics profiled cohort includes healthy controls and patients with** 87 **rheumatoid arthritis or osteoarthritis**

88 To characterize the gut-joint axis across successive stages of arthritis, we analyzed gut
89 metagenomic data and serum metabolomic data from 27 healthy controls, 76
90 rheumatoid arthritis (RA) patients, 19 osteoarthritis (OA) patients. RA samples were
91 grouped into four RA stages including RA stage 1 (RAS1, n = 15), RA stage 2 (RAS2,
92 n = 21), RA stage 3 (RAS3, n = 18), and RA stage 4 (RAS4, n = 22) according to the
93 rheumatoid diagnostic score¹⁵ (**Supplementary Tables 1**). We employed two separate
94 processes for gut metagenomic data analysis: 1) One was based on 4 million predicted
95 genes and investigated the stage-specific functional composition in arthritis
96 progression, as well as the metagenomic species (MGS) that most drove the
97 correlation of these microbial functions with arthritis¹⁶ (**Fig. 1a**); and 2) The other
98 reported the 232 classified microbial species composition across arthritis stages¹⁷,
99 profiled by MetaPhlAn2¹⁸ (**Fig. 1b**). We also explored the variance in serum
100 metabolomic across RA stages (**Fig. 1c**). Furthermore, we identified the most notable
101 variation of features at the multi-omics level that could identify each arthritic stage
102 from the healthy condition with high accuracy using the random forest (RF) algorithm
103 on microbial species, microbial KOs, and serum metabolites, respectively (**Fig. 1d**).
104 Finally, we attempted to determine the existence of microbial invasion of the joint in
105 the fourth stage of RA.

106 Patients with rheumatoid arthritis or osteoarthritis have altered microbial 107 composition compared to healthy controls

108 We firstly identified arthritis-related microbial species that were substantially elevated
109 or depleted in RA or OA groups, when compared with the healthy group ($P < 0.05$,
110 $FDR < 0.05$, Mann-Whitney-Wilcoxon test, **Supplementary Tables 2**). The union set
111 of these arthritis-related species ($n = 15$) was then used for linear discriminant
112 analysis (LDA) to separate arthritis samples from healthy samples (**Fig. 2a**). We
113 discovered that arthritis gut microbiota could be discriminated from healthy
114 microbiota, while we did not find an evident separation between RA and OA samples
115 in LDA. The most abundant two arthritis-related species were *Bacteroides uniformis*
116 whose proportion was mainly contributed by healthy samples, and *Escherichia. coli*
117 whose proportion was mainly contributed by RA and OA samples (**Fig. 2b**).
118 Moreover, the RA group contributed dominantly to the proportion of *Clostridium*
119 *hathewayi*, compared to its extremely low level in healthy and OA samples (**Fig. 2b,c**).
120 Thus, the significant elevation of *C. hathewayi* might be the RA-specific microbial
121 variation that could discriminate RA samples from OA samples.

122 Microbial species present stage-specific alterations throughout the progression of 123 rheumatoid arthritis

124 Next, we sought to explore the stage-specific taxonomic composition throughout the
125 RA progression. We examined 232 classified species and compared their abundances
126 in each of RA stages and OA to those in the healthy state (**Supplementary Tables 3**).
127 The elevated species in RA progression were mostly from the phyla Firmicutes and
128 Actinobacteria, while the depleted species were predominantly from the phylum
129 Bacteroides (**Fig. 2d**). Additionally, we exhibited the stage-specific abundance of the
130 identified 15 arthritis-related species, as well as species who or whose genera were
131 previously recognized as arthritis-related in other studies¹⁹ (**Fig. 2d**).

132 For the arthritis-depleted species (**Fig. 2d**), we found that *B. uniformis* (RAS1: $P =$
133 1.4×10^{-3} , RAS2: $P = 5.0 \times 10^{-3}$, RAS3: $P = 7.2 \times 10^{-3}$, RAS4: $P = 5.7 \times 10^{-3}$) and *C.*
134 *fastidiosus* (RAS1: $P = 0.041$, RAS2: $P = 0.024$, RAS3: $P = 0.022$) were significantly
135 depleted only during the RA progression but not in OA, whereas *B. cellulosilyticus*
136 was significantly depleted in OA ($P = 4.72 \times 10^{-4}$), and in the very early and late
137 stages of the RA progression (RAS1: $P = 0.023$, RAS4: $P = 6.24 \times 10^{-3}$).

138 For the arthritis-elevated species (**Fig. 2d**), we found that *C. hathewayi* was not
139 significantly elevated in the very early RA stage, though kept a significantly higher
140 level during the subsequent RA stages than that in the healthy state (RAS2: $P = 2.52 \times$
141 10^{-3} , RAS3: $P = 5.92 \times 10^{-3}$, RAS4: $P = 2.96 \times 10^{-3}$). In addition, *Bifidobacterium*
142 *dentium* also maintained considerable increased abundance from RAS2 to RAS4
143 (RAS2: $P = 7.16 \times 10^{-3}$, RAS3: $P = 3.70 \times 10^{-3}$, RAS4: $P = 9.15 \times 10^{-4}$). *B. dentium*
144 was reported to be associated with the development of dental caries and periodontal
145 disease, both of which were particularly prevalent in patients with RA^{20,21}.
146 *Streptococcus anginosus* and *S. parasanguinis* kept significant elevation ($P < 0.05$)
147 throughout the RA progression (**Fig. 2d**), while a few of other species from genus
148 *Streptococcus* were elevated stage-specifically. *S. mitis* ($P = 1.21 \times 10^{-3}$) and *S.*
149 *sanguinis* ($P = 5.37 \times 10^{-4}$) were mainly elevated in RAS1, while *S. salivarius* was
150 mainly elevated in RAS3 ($P = 0.035$) and RAS4 ($P = 0.022$). These results suggest
151 that although a few microorganisms were identified as correlated with arthritis, they
152 might not all keep abundant or depleted throughout the progression. Part of them only
153 preferred certain stages.

154 **Certain microbial species are only significantly altered in a single stage of** 155 **rheumatoid arthritis**

156 Besides those varied species shared by two or more arthritis stages, we also cared
157 about those species that were only elevated or depleted at one specific arthritis stage
158 when compared to the healthy state ($P < 0.05$, Mann-Whitney-Wilcoxon test,
159 **Supplementary Tables 4**). *Collinsella aerofaciens*, for example, was found

160 significantly elevated only in RAS1 ($P = 0.043$). *C. aerofaciens* was previously
161 reported to generate severe arthritis when inoculated into collagen-induced arthritis
162 (CIA)-susceptible mice, and an *in vitro* experiment showed that *C. aerofaciens* could
163 increase gut permeability and induce IL-17A expression, a key cytokine involved in
164 RA pathogenesis¹¹. Therefore, the significant elevation of *C. aerofaciens* might
165 contribute to the early breach in gut barrier integrity during RA progression⁵. The
166 translocation of microbial products through the breach would then trigger the
167 subsequent clinical arthritis⁵.

168 *Veillonella parvula*, whose infection could cause osteomyelitis²², was found elevated
169 exclusively in RAS3 ($P = 0.027$). *Eggerthella lenta* ($P = 0.018$) and *Bifidobacterium*
170 *longum* ($P = 0.022$) were found elevated exclusively in RAS4. Additionally, mice
171 gavaged with *E. lenta* were reported to show an increased gut permeability and
172 produce proinflammatory cytokines IL-6, IL-21, IL-23²³. *Dialister invisus* was found
173 significantly elevated only in OA microbiota ($P = 0.041$). *D. invisus* was previously
174 reported to be more abundant in the inflamed ileal and colonic biopsy from patients
175 with spondyloarthritis than in the noninflamed biopsy, and was positively correlated
176 with spondyloarthritis severity²⁴. We also found that the species *Bacteroides ovatus*,
177 *Bacteroides coprocola*, and *Bacteroides faecis* were depleted significantly only in
178 RAS1 ($P = 4.2 \times 10^{-3}$), RAS2 ($P = 0.036$), and RAS4 ($P = 0.020$), respectively. These
179 results suggest that certain microbial species in arthritis patients present more
180 abundant or more depleted exclusively in a particular arthritis stage, for implementing
181 specific functions such as increasing gut permeability and inducing proinflammatory
182 cytokines.

183 ***B. uniformis*- and *B. plebeius*-driven microbial dysfunction in glycosaminoglycan** 184 **metabolism promotes arthritis progression by hurting articular cartilage**

185 To understand the variation of microbial functions across arthritis stages, we grouped
186 4,047,645 metagenomic genes into 6224 KEGG orthologs (KOs) and 404 KEGG
187 modules. We identified 12 KEGG modules that were significantly correlated with

188 arthritis ($P < 0.05$, FDR < 0.1 , Mann-Whitney-Wilcoxon test, **Fig. 3a**, **Extended**
 189 **Data Fig. 1**, **Supplementary Tables 5** and **6**). We then compared KOs' abundances
 190 within these arthritis-related KEGG modules between arthritis stages and healthy state
 191 (**Fig. 3a** and **Supplementary Tables 7**), and used leave-one-out analysis to identify
 192 the metagenomic species (MGS) that most drove the correlations of these KEGG
 193 modules with arthritis (**Fig. 3b** and **Extended Data Fig. 2**).

194 We found an evident decrease in glycosaminoglycan (CAG) metabolism in all the
 195 arthritis stages. It was mainly reflected by the significant decrease in K01197
 196 (hyaluronoglucosaminidase, RAS1: $P = 2.31 \times 10^{-3}$, RAS2: $P = 7.78 \times 10^{-5}$, RAS3: P
 197 $= 2.00 \times 10^{-5}$, RAS4: $P = 6.70 \times 10^{-4}$, OA: 8.65×10^{-4}) of dermatan sulfate (DS)
 198 degradation and the significant decrease in K10532 (heparan-alpha-glucosaminide
 199 N-acetyltransferase, RAS1: $P = 2.30 \times 10^{-4}$, RAS2: $P = 0.045$, RAS3: $P = 3.2 \times 10^{-3}$,
 200 RAS4: $P = 4.66 \times 10^{-3}$, OA: $P = 3.72 \times 10^{-4}$) of heparan sulfate (HS) degradation
 201 (**Fig. 3a**). The compressive features of the articular cartilage are determined by CAGs,
 202 predominantly chondroitin sulfate (CS) and hyaluronic acid²⁵. CS is a major
 203 component of the extracellular matrix of many connective tissues, including cartilage,
 204 bone, skin, ligaments, and tendons²⁶. We found that the significant depletion of DS
 205 degradation at all the arthritis stages would decrease the production of chondroitin
 206 4-sulfate (**Fig. 3b**), which hurt the mechanical properties of articular cartilage of
 207 patients with arthritis. Additionally, total CAGs, CS, and heparan sulfate/heparin
 208 (HS/H) were reported to be higher in plasma of RA patients than in plasma of healthy
 209 subjects²⁷. Moreover, the sulfation of HS was reported to be increased in cartilages of
 210 patients with OA²⁸, and a decreased HS content or a reduced sulfation level was
 211 reported to protect against OA progression by regulating protease activity²⁵. Thus, the
 212 significant depletion of HS degradation in microbial function would prevent HS from
 213 being degraded, promoting arthritis progression (**Fig. 3b**). The most driving species of
 214 dermatan sulfate degradation and heparan sulfate degradation were MGS *Bacteroides*
 215 *uniformis* and MGS *Bacteroides plebeius*, respectively. The genes of MGS *B.*
 216 *uniformis* related to K01197 were found most depleted in RAS2, while the genes of

217 MGS *B. plebeius* related to K10532 were found most depleted in RAS3 and RAS4
218 (**Fig. 3B**). These results indicated that the depleted microbial function in DS
219 degradation and HS degradation driven by *B. uniformis* and *B. plebeius*, respectively,
220 could promote RA and OA in a way of hurting articular cartilage.

221 ***E. coli*-driven microbial dysfunction promotes inflammation in arthritis**

222 **pathogenesis particularly in the second and third stages of RA**

223 We also identified certain arthritis-elevated microbial functions that were related to
224 pro-inflammation. As we have previously reported, *Escherichia coli* could promote
225 arthritis progression by enhancing ascorbate degradation⁹. Ascorbate could prevent
226 the development of inflammatory arthritis by facilitating collagen synthesis,
227 moderating autoimmune responses, and ameliorating inflammation^{29,30}. Here, we
228 found most of the KOs related to ascorbate degradation kept a higher level in RA and
229 OA progression, especially in RAS2 and RAS3 (**Fig. 3a**). MGS *E. coli* drove the
230 correlations of ascorbate degradation with arthritis. Genes of K02821
231 (phosphotransferase system) in RAS1, K03475 (phosphotransferase system), K03476
232 (L-ascorbate 6-phosphate lactonase), and K03479 (L-ribulose-5-phosphate
233 3-epimerase) in all arthritis stages were mostly contributed by MGS *E. coli* (**Fig. 3b**).
234 Hence, the deficiency of ascorbate reported in patients with RA³¹, might be related to
235 the enhanced ascorbate degradation in gut microbial dysfunction. Moreover, we found
236 that the enhanced ascorbate degradation was evidently correlated with the elevation of
237 multi serum cytokines, including IL-1 β ($P = 5.44 \times 10^{-4}$, $q = 0.037$), IL-17 ($P = 6.59$
238 $\times 10^{-4}$, $q = 0.050$), IL-8 ($P = 7.45 \times 10^{-4}$, $q = 0.07$), IL-2 ($P = 9.37 \times 10^{-4}$, $q = 0.064$),
239 IL-6 ($P = 1.12 \times 10^{-3}$, $q = 0.061$), IL-12p70 ($P = 3.47 \times 10^{-3}$, $q = 0.093$), IL-4 ($P =$
240 4.2×10^{-3} , $q = 0.097$), TNF- α ($P = 6.59 \times 10^{-3}$, $q = 0.050$).

241 To confirm the effects of ascorbate on RA progression, we examined the serum
242 TNF- α level and IL-6 level, bone CT scans, and bone density of 1) normal DBA/1
243 mice, 2) DBA/1 mice with CIA, and 3) DBA/1 mice with CIA and gavage of
244 ascorbate. We found that the three-month gavage of ascorbate can prevent the

245 increase of TNF- α and IL-6 levels, inhibit bone destruction and bone loss (**Extended**
246 **Data Fig. 3**).

247 For other arthritis-elevated microbial functions, the arginine succinyltransferase
248 pathway was found significantly elevated mainly in RAS2 and RAS3 (**Fig. 3a**).
249 L-arginine has been reported to be able to prevent bone loss induced by zinc oxide
250 nanoparticles or by cyclosporin A, through anti-inflammatory mechanism³² or nitric
251 oxide production, respectively³³. The most driven MGS of the arginine
252 succinyltransferase pathway was MGS *E. coli* (**Fig. 3b**). Moreover, another KEGG
253 module named trans-cinnamate degradation was also most driven by MGS *E. coli*,
254 whose most KOs were significantly elevated in RA and OA, especially in RAS2 (**Fig.**
255 **3a,b**). Trans-cinnamate degradation was positively correlated with the elevation of
256 multi serum cytokines, including IL13 ($P = 1.63 \times 10^{-5}$, $q = 4.41 \times 10^{-3}$), IL-1 β ($P =$
257 2.87×10^{-5} , $q = 3.90 \times 10^{-3}$), IL10 ($P = 4.10 \times 10^{-5}$, $q = 5.57 \times 10^{-3}$), TNF- α ($P =$
258 6.97×10^{-5} , $q = 9.47 \times 10^{-3}$), IL2 ($P = 7.40 \times 10^{-5}$, $q = 0.010$), IL-8 ($P = 2.46 \times 10^{-4}$,
259 $q = 0.067$), IFN γ ($P = 3.92 \times 10^{-4}$, $q = 0.036$), IL4 ($P = 1.78 \times 10^{-3}$, $q = 0.061$), and
260 IL-12p70 ($P = 2.61 \times 10^{-3}$, $q = 0.089$). Taken together, these results suggested that
261 arthritis-related microbial functional change could promote RA and OA progression
262 by hurting bone tissue and strengthening inflammation. The inflammation-related
263 microbial function was extremely active in RAS2 and RAS3 and largely driven by *E.*
264 *coli*.

265 **Gut microorganisms may invade joint synovial fluid in the fourth stage of RA**

266 To examine whether and in which of stages microorganisms had invaded joint
267 synovial fluid (SF) of patients with RA, we collected samples of joint SF from
268 another cohort of 271 patients in four distinct stages of RA, including RAS1 ($n = 52$),
269 RAS2 ($n = 66$), RAS3 ($n = 67$), and RAS4 ($n = 86$). Enhanced gut permeability may
270 render it possible for microorganisms and their products to translocate, triggering an
271 immune response^{5,34}. Additionally, a few studies have shown the presence of
272 microbial DNA in the joint SF³⁵⁻³⁸. We thus speculated that gut microorganisms might

invade the joint SF of patients with RA through the gut-joint axis. To test this speculation, we next perform 16S rRNA gene sequencing on the SF samples. Notably, we were not able to obtain enough bacterial DNA for sequencing in samples of RAS1, RAS2, or RAS3, however, we could identify many microorganisms in samples of joint SF from patients in RAS4 (**Extended Data Fig. 4**). We found that most of the joint-SF microorganisms were from phyla Proteobacteria and Firmicutes (**Extended Data Fig. 4a**). Moreover, we could recognize *Eggerthella lenta* and *Bifidobacterium longum* in most of the SF samples (**Extended Data Fig. 4b**), both of which were observed to be exclusively elevated in the intestine of patients in RAS4 (**Supplementary Tables 4**). In addition, *Prevotella copri* that has been reported highly correlated with RA^{10,39} was also found abundant in most SF samples of patients in RAS4.

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

Rheumatoid arthritis has stage-specific biomarkers indicating both microbial dysbiosis and metabolic disorders

Microbial dysbiosis has exhibited stage-specific patterns, playing a crucial role in successive stages of RA, via microbial dysfunction. It was also speculated that dysbiosis might promote RA progression via metabolic perturbation⁵.

To determine the most discriminant characteristics of each arthritis stage at the multi-omics level, we performed an RF algorithm on 232 microbiome species, 6224 KOs, and 277 metabolites (**Fig. 4**), and displayed the most notable microbial and metabolic features of each stage in **Fig. 5a**. Metabolites exhibited the best AUROC in discriminating samples of all the arthritis stages from healthy samples, with AUROC ranging from 0.974 to 0.998. Other characteristics at the species and KO levels exhibited weaker discriminant ability than features at the metabolite level, with AUROC ranging from 0.760 to 0.838 and 0.799 to 0.852, respectively (**Fig. 4**). Thus, metabolic disorders were predominant in each stage of arthritis, implicating a critical involvement in pathogenesis throughout the disease development.

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

Each of the arthritis stages had the respective most prominent change in metabolites (**Fig. 4** and **Fig. 5a**), even though all of them were strongly different from those of health status (**Fig. 5b**). The most prominent changes in metabolites were the significant increase of DL-lactate and gly-glu in RAS1 ($P = 2.15 \times 10^{-6}$, $P = 4.70 \times 10^{-4}$), the decrease of N,N-dimethylaniline and the increase of methoxyacetic acid in RAS2 ($P = 4.60 \times 10^{-8}$, $P = 1.28 \times 10^{-8}$), the increase of cysteine-S-sulfate ($P = 4.66 \times 10^{-12}$) in RAS3, the increase of galactinol and 3 α -mannobiose in RAS4 ($P = 5.71 \times 10^{-5}$, $P = 5.68 \times 10^{-4}$), and the decrease of N,N-dimethylaniline and increase of gly-glu ($P = 2.00 \times 10^{-7}$, $P = 1.74 \times 10^{-3}$) in OA, as compared to a healthy state. A considerable increase in lactate, which was the most noticeable alteration in metabolites in RAS1 (**Fig. 4a**), occurred early in RA development and remained throughout the subsequent RA stages and OA (**Fig. 5b**). Lactate in the inflamed tissue was reported to promote the upregulation of the lactate transporter SLC5A12 by human CD4⁺ T cells⁴⁰. This upregulation led to increased IL17 production via nuclear PKM2/STAT3 and enhanced fatty acid synthesis⁴⁰. In addition, antibody-mediated blockade of SLC5A12 could ameliorate the severity of arthritis in a mouse model⁴⁰. We found serum DL-lactate was positively correlated with a certain number of

inflammatory cytokines such as IL-6 ($r = 0.33$, $P = 1.86 \times 10^{-5}$, $q = 3.69 \times 10^{-4}$), IL-17 ($r = 0.32$, $P = 4.23 \times 10^{-5}$, $q = 9.76 \times 10^{-4}$) and TNF-alpha ($r = 0.40$, $P = 1.84 \times 10^{-7}$, $q = 1.28 \times 10^{-5}$). As the end product of glycolysis⁴¹, the lactate production would be affected by the intervention of glycolysis, such as the disruption of the glucose-6-phosphate isomerase (GPI, K06859) and phosphofructokinase (PFK, K00850). Both GPI and PFK, as the upstream glycolytic enzyme, were found significantly decreased in RA stages ($P < 0.05$, $q < 0.1$), while GPI was not significantly decreased in the OA stage (**Fig. 3a**). As a consequence, microbial functional variation in glycolysis might result in an increased concentration of DL-lactate at RA stages compared to OA stages and healthy state, contributing to the elevated inflammatory cytokines in RA stages.

Metabolic disorders break the balance between osteoblasts and osteoclasts in the medium stage of rheumatoid arthritis

Furthermore, we discovered that a few noticeable alterations in metabolites might distinguish a given arthritis stage from not just a healthy state but also other arthritis stages. (**Fig. 5b**): (1) DL-lactate in OA was less than that at all RA stages ($P = 0.037$); (2) Methoxyacetic acid in RAS2 ($P = 1.68 \times 10^{-4}$) or cysteine-S-sulfate in RAS3 ($P = 2.42 \times 10^{-4}$) or Galactinol and 3 α -mannobiose in RAS4 ($P = 9.37 \times 10^{-3}$, $P = 4.89 \times 10^{-3}$), respectively, was higher than that at all the other RA stages and OA. Such dramatic changes in arthritis serum metabolites suggested that they might be more than just a metabolomic depiction of each stage of arthritis; they could also be potential promoters for the next stage. For instance, methoxyacetic acid was reported to have inhibitory effects on osteoblasts and could cause reductions in bone marrow cellularity⁴²⁻⁴⁴. The dominant elevation of methoxyacetic acid in RAS2 might be a precursor to bone erosion in RAS3. Moreover, the notable elevation of cysteine-S-sulfate in RAS3 might promote the osteoclasts that occurred in the late RA stages. cysteine-S-sulfate, a structural analog of glutamate, acted as an agonist of N-methyl-D-aspartate receptor (NMDA-R) whose expression and function in osteoclasts engaged in bone resorption⁴⁵. Taken together, aberrant elevations of

methoxyacetic acid in RAS2 hindered osteoblasts, whereas abnormal elevations of cysteine-S-sulfate in RAS3 encouraged osteoclasts. As a result of the imbalance between osteoblasts and osteoclasts, bone degradation occurred in the late stages of RA.

Discussion

We performed metagenomic and metabolomic studies on a cohort of 122 RA patients, together with 16S rRNA gene sequencing and scanning electron microscopy on joint synovial fluid samples from another cohort of 271 RA patients, and drew several important conclusions. First, we found that microbial dysbiosis and metabolic disorders presented dynamic patterns across successive stages of RA. Second, these dysbiosis and disorders continuously promoted RA progression by increasing gut permeability, inducing inflammation, and hurting bone tissue. Microbial dysfunction in dermatan sulfate and heparan sulfate degradation continuously hurting the articular cartilage across successive stages of RA, as well as in OA. Third, certain of these influences were more concentrated on a particular stage of RA. Microbial dysfunction related to inflammation was focused on the second and third stages and was mostly driven by *E. coli*. Metabolic disorders induced bone erosion through the abnormal elevation of methoxyacetic acid in the second stage and cysteine-S-sulfate in the third stage. Finally, there was strong evidence that microbial invasion of the joint happened in the fourth stage of RA. Synovial-fluid 16S rRNA gene sequencing was capable of detecting *Eggerthella lenta* and *Bifidobacterium longum*, both of which were elevated in fecal metagenomics of patients, exclusively in the fourth stage of RA rather than other stages. Moreover, *Clostridium sp.*, *Enterococcus sp.*, and *Citrobacter sp.* were separated from the joint synovial fluid. Under scanning electron microscopy, we could observe substances shaped like bacteria in rod-like or spherical forms. However, we were unable to detect bacteria in the joint synovial fluid from patients in other stages of RA due to the low DNA content. We conclude that the buildup of microbial dysbiosis and metabolic disorders across successive RA stages may induce the microbial invasion of the joint in the last stage of RA.

387 The findings of the continuous effects of microbial dysbiosis and metabolic disorders
 388 across successive RA stages could be of considerable clinical importance (**Fig. 6**).
 389 First, hurting bone tissue (**Fig. 6a**). Depleted *B. uniformis* drove the decreased
 390 dermatan sulfate degradation, leading to less production of chondroitin 4-sulfate. The
 391 low content of chondroitin 4-sulfate would inhibit the formation of articular
 392 cartilage^{25,26}. Additionally, depleted *B. plebeius* drove the less degradation of HS. A
 393 high HS content would promote OA progression^{25,28}. Second, increasing
 394 inflammation particularly in the second and the third of RA (**Fig. 6b**). Elevated *E. coli*
 395 drove enhanced degradation of L-Arginine through the arginine succinyltransferase
 396 pathway. L-Arginine acted as an inhibitor of bone loss via anti-inflammation³² and
 397 nitric oxide production³³. Besides, elevated *E. coli* drove enhanced ascorbate
 398 degradation. Ascorbate is an anti-inflammatory vitamin^{29,30}. We found that three
 399 months of ascorbate treatment to mice with CIA could relieve arthritis, by preventing
 400 bone destruction and inflammation. Owing to the sequencing limitation in this study,
 401 we are unable to pinpoint the strain of microbial species in these gut-joint axes.
 402 Further explorations are required for determining strain identity, maybe through
 403 third-generation sequencing, as well as for testing its effects in a sterile mice model.
 404 In clinical practice, the intervention based on these proposed mechanisms may
 405 profoundly improve the therapy for patients during the whole RA progression,
 406 through restoring *B. uniformis* and *B. plebeius*, and inhibiting *E. coli*, or supplying
 407 patients with ascorbate and chondroitin 4-sulfate.

408 Apart from clinical treatments for microbial dysbiosis and metabolic disorders in the
 409 whole RA progression, specific treatments are necessary for each RA stage to not
 410 only relieve the symptoms in the current stage but also inhibit the onset of the
 411 subsequent stages. Our findings suggested that certain effects of microbial dysbiosis
 412 or metabolic disorders on RA progression were more focused on a particular stage. In
 413 the early stage of RA, elevated *C. aerofaciens* and depleted *Parabacteroides*.
 414 *distasonis* both contributed to the increased gut permeability (**Fig. 6b**). *C. aerofaciens*
 415 was found elevated exclusively in the first stage of RA, rather than other stages. *C.*

416 *aerofaciens* was reported to induce expression of IL-17 network cytokines and
 417 increase gut permeability via reducing the expression of tight junction protein in
 418 epithelial cells¹¹. *P. distasonis* was depleted in most of RA stages and OA with a
 419 remarkable ability to distinguish arthritis samples from healthy samples, particularly
 420 in the first stage (**Fig. 5b**). Previous studies have reported the anti-inflammation effect
 421 of *P. distasonis* such as its correlation with CRP⁴⁶, TNF- α ⁴⁷, and IL-1 β ⁴⁷, and that the
 422 oral administration of its antigens could reduce intestinal inflammation in colitis
 423 mice⁴⁸. However, the role of *P. distasonis* in RA pathogenesis remained unknown. A
 424 recent study found that *P. distasonis* could preserve the intestinal barrier integrity in
 425 azoxymethane-treated A/J mice, and possess anti-inflammatory properties and
 426 anti-tumor properties through the inhibition of signaling via TLR4, MYD88, and Akt,
 427 as well as the stimulation of apoptosis⁴⁹. We thus speculate that the considerable
 428 depletion of *P. distasonis* occurs in the very early of RA, contributing to inflammation
 429 and the collapse of intestinal barrier integrity, though this possible mechanism
 430 warrants further investigations. Intervention on *C. aerofaciens* and *P. distasonis*
 431 therefore may have a higher priority in the first stage of RA than in subsequent stages
 432 to inhibit the early increase of gut permeability and inflammation.

433 In the second stage of RA, the aberrant elevation of methoxyacetic acid had inhibitory
 434 effects on osteoblasts and caused reductions in bone marrow cellularity⁴²⁻⁴⁴ (**Fig. 6b**).
 435 The inhibited osteoblasts drew the foreshadowing for the bone erosion that happened
 436 in the next stage. In the third stage, the considerable elevation of cysteine-S-sulfate
 437 enhanced the osteoclasts by NMDA-R interaction⁴⁵ (**Fig. 6b**). The imbalance between
 438 osteoblasts and osteoclasts then resulted in bone erosion that happened in the third
 439 stage and persisted in the late RA stages (**Fig. 6b**). Thus, methoxyacetic acid may be a
 440 targeted metabolite for treatment to patients in the second stage of RA and serve as a
 441 precaution against the upcoming third stage. In the fourth stage of RA, continuous
 442 damages in gut permeability, intestinal immunity, and bone tissue resulted in the
 443 extremely high elevation of proinflammatory cytokines and severe bone erosion, and
 444 even microbial invasion of the joint. Due to the collapse of the intestinal barrier,

445 microbes and microbial metabolites might be transmitted to the joint via blood⁵. In
446 addition to routine medical therapies for immunity regulation, specific treatments to
447 the microbes in the joint synovial fluid of the patients in the fourth stage of RA may
448 ameliorate the joint micro-environment to decrease synovial inflammation and
449 prevent potential bacterial effects.

450 This study also has prospects and limitations. Firstly, it may aid in enhancing RA
451 therapy across populations by assessing whether RA patients with different genetic
452 backgrounds exhibit distinct gut-joint links across stages. Secondly, a long-term
453 follow-up investigation on a single individual throughout his/her RA development
454 may reinforce the conclusions of this study. Lastly, additional researches into the
455 joint-SF microbiome and metabolome have the potential to figure out more
456 sophisticated mechanisms underlying RA pathogenesis.

457 In conclusion, we demonstrate dynamic shifts in gut microbiome and serum
458 metabolome, and their continuous roles in pathogenesis of RA across four successive
459 stages. Microbial dysbiosis and metabolic disorders promotes RA through increasing
460 gut permeability, triggering inflammation, and hurting bone tissue. Although these
461 effects persist in the whole RA progression, certain of them are more focused on
462 specific stages. Furthermore, the buildup of the effects of these dysbiosis and
463 disorders may lead to the collapse in the fourth stage of RA, as represented by the
464 microbial invasion of the joint. This study highlights that different RA stages
465 corresponds to different gut-joint axes that may contribute to different symptoms of
466 patients. The stage-specific intervention in microbial dysbiosis and metabolic
467 disorders was warranted for prognosis and prevention of RA progression.

468 **Methods**

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

470 A total of 122 fecal and 122 serum samples were collected from 122 outpatients of the
471 Shandong Provincial Qianfoshan Hospital (Jinan, Shandong, China). These

472 outpatients included 27 healthy individuals, 19 patients with osteoarthritis (OA), and
 473 76 patients with rheumatoid arthritis (RA). RA patients were grouped into four RA
 474 stages including RAS1 (n = 15), RAS2 (n = 21), RAS3 (n = 18), and RAS4 (n = 22)
 475 according to the rheumatoid diagnostic score¹⁵, where RAS1, RAS2, RAS3, and
 476 RAS4 has a score of 6-7, 8, 9, and 10, respectively (**Supplementary Tables 1**). The
 477 score was evaluated by the sum of four categories¹⁵: A) joint involvement that refers
 478 to any swollen or tender joint on examination; B) Serology determined by values of
 479 rheumatoid factor and anti-citrullinated protein antibody; C) Acute-phase reactants
 480 determined by values of C-reactive protein and erythrocyte sedimentation rate; and D)
 481 Duration of symptoms. Subsequently, fecal samples were sequenced and serum
 482 samples were used to test the serum metabolites and inflammatory cytokine levels.
 483 Serum metabolites were tested using UHPLC-MS/MS, while inflammatory cytokines
 484 were quantified using the MESO SCALE DISCOVERY (MSD) Quick Plex S600MM
 485 multiplex assay. Cytokine levels of healthy individuals were not available because of
 486 their extremely low levels. All of the participants were at fasting status during the
 487 sample collection in the morning. Only participants who met the standard were
 488 recruited in this study: Recruited individuals had not received treatment in the recent
 489 month, and had no malignant tumor, no other rheumatic diseases such as ankylosing
 490 spondylitis, psoriasis, gout, no gastrointestinal diseases such as diarrhea, constipation,
 491 and hematochezia in the recent one month, no infections, no other comorbidity such
 492 as diabetes and hepatitis B.

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

494 A total of 271 patients with RA of four distinct stages were recruited, including 52
 495 RAS1, 66 RAS2, 67 RAS3, and 86 RAS4. RA patients were grouped into four RA
 496 stages according to the rheumatoid diagnostic score¹⁵. The patients met the
 497 aforementioned standard. Synovial fluid (SF) samples were collected aseptically from
 498 knee joints during therapeutic aspiration. SF samples were deposited in sterile tubes
 499 on ice and homogenized within five minutes of collection. A tube filled with sterile
 500 phosphate-buffered saline (PBS) was left open throughout the procedure and

501 subsequently processed in parallel with the samples as a negative control. The entire
502 experiment was conducted in a completely sterile atmosphere. Each sample was
503 immediately frozen and kept without heparin or hyaluronidase at -135°C . For each
504 patient, a total of 7ml SF was collected, of which 5ml was utilized for 16S rRNA gene
505 sequencing, 1ml was used for bacteria isolation, and 1ml SF was prepared for
506 scanning electron microscopy.

507 **Metagenome sequencing and processing**

508 Whole-genome shotgun sequencing of fecal samples was carried out on the Illumina
509 HiSeq X Ten. All samples were paired-end sequenced with a 150-bp read length.
510 Reads containing more than 40 base pairs with low-quality values of 38 or less were
511 discarded. Reads containing more than 10 unidentified base pairs were discarded.
512 Reads containing more than 15 base pairs that overlapped adapter sequences were
513 discarded. Next, reads were mapped against the human genome using Bowtie2
514 (version 2.2.9). Those that were mapped were considered to be derived from the
515 human genome and were discarded. Finally, unpaired reads were discarded.
516 High-quality reads were used for the following analyses. MetaPhlAn2¹⁸ was then used
517 to generate microbial species profiles.

518 **Contig assembly and ORF prediction**

519 The reads were assembled into contigs using MEGAHIT¹⁷ (version 1.2.6) with the
520 minimum contig length set at 500 bp and default parameters for metagenomics. The
521 open reading frames (ORFs) were predicted from the assembled contigs using
522 Prodigal⁵⁰ (version 2.6.3) with default parameters. The ORFs of <100 bp were
523 removed.

524 **Non-redundant gene catalogue construction**

525 The ORFs were clustered to remove redundancy using Cd-hit⁵¹ (version 4.6.6) with a
526 sequence identity threshold set at 0.95 and the alignment coverage set at 0.9, which

527 resulted in a catalogue of 4,047,645 nonredundant genes. The non-redundant genes
528 were then collapsed into metagenomic species (MGS)^{16,52}, and grouped into KEGG
529 functional modules¹⁶, as described below.

530 **Identification of metagenomic species**

531 High-quality reads were mapped to the catalogue of nonredundant genes using Bowtie
532 2⁵³ (version 2.2.9) with default parameters. The abundance profile for each catalogue
533 gene was calculated as the sum of uniquely mapped sequence reads, using 19M
534 sequence reads per sample (downsized). The co-abundance clustering of the
535 4,047,645 genes was performed using canopy algorithm⁵²
536 (<http://git.dworzynski.eu/mgs-canopy-algorithm>), and 553 gene clusters that met the
537 previously described criteria⁵² and contained more than 700 genes were referred to as
538 MGS. MGS that were present in at least 4 samples were used for the following
539 analysis. The abundance profiles of MGS were determined as the medium gene
540 abundance throughout the samples. MGS were taxonomically annotated by summing
541 up the taxonomical annotation of their genes as described by Nielsen *et al*⁵². Each
542 MGS gene was annotated by sequence similarity to the bacterial complete genome in
543 NCBI Reference Sequence Database (BLASTN, E-value < 0.001).

544 **Annotation of KEGG modules.**

545 The catalogue of the nonredundant genes was functionally annotated to the KEGG
546 database (release 94.0) by KofamKOALA (version 1.3.0)^{54,55}. The produced KEGG
547 Orthologs (KOs) were mapped to the KEGG modules annotation downloaded on 1
548 August 2020 from the KEGG BRITE database. KOs that were present in at least 4
549 samples were used for the following analysis. The KO abundance profile was
550 calculated by summing the abundance of genes that were annotated to each of the
551 KOs.

552 **Leave-one-out analysis**

553 Leave-one-out analysis was used to test which MGS was driving the observed
554 associations between KEGG modules and arthritis. The calculation of the KO
555 abundance was iterated excluding the genes from a different MGS, in each iteration.
556 The effect of a given MGS on a specified association was defined as the change in
557 median Spearman correlation coefficient between KOs and arthritis when genes from
558 the respective MGS were left out, as previously described^{16,56}.

559 **Multi-omics biomarkers**

560 A random forest algorithm was performed using the R package “randomForest” on
561 each of the three profiles independently, including 232 microbiome species, 6,224
562 KOs, and 277 metabolites. Function “trainControl” in R package “caret” was used to
563 perform 10 repeats of 10-fold cross-validation for each data set. Function “train” in R
564 package “caret” was used to fit models over different tuning parameters to determine
565 the “mtry” for random forest algorithm.

566 **Correlation of phenotypes and KEGG modules**

567 The clinical phenotypes including types of arthritis (Healthy = 0, OA = 1, RA = 2)
568 were used in the correlation analysis with KEGG modules. The associations of
569 clinical phenotypes with KEGG modules were determined by if Spearman
570 correlations of the phenotype with the abundances of KOs in the given KEGG module
571 were significantly higher or lower (Mann–Whitney U-test FDR < 0.1) than with the
572 abundances of all other KOs. The phenotypes adjusted by age and gender were also
573 tested using partial Spearman correlation. The union set of the significant associations
574 between KEGG modules and phenotypes/phenotypes adjusted by age and gender
575 were used for the following analysis.

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

577 Nine healthy seven-week-old DBA/1 mice weighing 20g were fed in an ultra-clean
578 animal laboratory (SPF grade) with a humidity of 55% and a temperature of 26°C.
579 CIA models were constructed and established as described before⁵⁷. Nine mice were
580 then divided into three groups (three mice per group), including normal DBA/1 mice
581 and two groups of DBA/1 mice with CIA. Three-month gavage (0.3ml/d) to three
582 groups of mice was conducted, including 1) 0.9% normal saline to normal DBA/1
583 mice, 2) 0.9% normal saline to DBA/1 mice with CIA, and 3) 100ng/ul ascorbate to
584 DBA/1 mice with CIA. After three-month gavage, the mice serum TNF- α level and
585 the IL-6 level were tested using ELISA kit (mlbio, China). Mice were then killed and
586 preserved in 4% formalin for two days. Micro-CT (QuantumGX, PerkinElmer,
587 UnitedStates) was used to perform scanning and three-dimensional structural
588 reconstruction of the joints. The settings were set to 209m, 90kV X-ray tube voltage,
589 160uA of the current, and 3 minutes of the scan time. The angle of the X-ray scan
590 rotated 180 degrees.

591 **16S rRNA gene sequencing and processing**

592 Bacterial DNA was extracted from 271 5ml synovial fluid samples. The tube
593 containing PBS serves as environmental control. 86 SF samples from patients in
594 RAS4 had bacteria DNA content (≥ 10 ng) (Bacterial DNA Kit, TIANGEN) for
595 bacteria 16S rRNA gene high-throughput sequencing. The V1/V2 hypervariable
596 regions of the 16S ribosomal RNA gene were sequenced using the Illumina HiSeq
597 platform. The 16S sequence paired-end data set was joined and quality filtered using
598 the FLASH as previously described⁵⁸. Sequence analysis was performed using the
599 Quantitative Insights into Microbial Ecology (QIIME, version 1.9.1) software⁵⁹.
600 Chimeric sequences were removed using de novo chimera detection of USEARCH
601 6.1⁶⁰. Sequences were clustered against the 2013 Greengenes (13_8 release)
602 ribosomal database's 97% reference data set. Sequences that did not match any entries
603 in this reference were subsequently clustered into de novo OTUs at 97% similarity

with UCLUST⁶⁰. Taxonomy was assigned to all OTUs using the RDP classifier⁶¹ against Greengenes reference data set.

Isolation of bacteria in the joint synovial fluid

Due to that pH of joint SF was 7.6, we used Luria-Bertani (LB) broth medium with a pH of 7.6 for cultivation. The procedures were as follows: 1) sterilized LB medium (pH7.6) and inoculation equipment were placed in an anaerobic glove box (Ruskinn Concept 400) for deaeration one day in advance; 2) 1ml SF samples per stage of RA were used for bacterial cultivation. 3) The SF sample was serially diluted with sterilized water, plated onto LB (0.5% yeast extract, 1% tryptone, 1% sodium chloride) agar medium, and then incubated at 37°C for 72 h to obtain single colonies. Then, three colonies per plate were randomly selected and streaked three consecutive times on LB agar medium to obtain a pure culture, which was named isolate SF1 to SF9, respectively. To identify isolate SF1 to SF9, a partial fragment of 16S rDNA was amplified with the primer pair 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3'), and then DNA sequencing was performed for preliminary identification.

Scanning electron microscopy for bacteria in the joint synovial fluid

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

630 **Statistical analyses**

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

638 **Data availability**

639 Whole-genome shotgun sequencing data are available in the Genome Sequence
 640 Archive (GSA) section of the National Genomics Data Center (project accession
 641 number CRA004348). 16S rRNA gene sequencing data are available in the Genome
 642 Sequence Archive (GSA) section of the National Genomics Data Center (project
 643 accession number CRA005811).

644 **Acknowledgements**

645 This work was partially supported by the National Natural Science Foundation of
 646 China (grant numbers 31871334, 82003766, 32071465, and 31671374) and the
 647 Academic Promotion Project of Shandong First Medical University (grant number
 648 2019LJ001).

649 **Author contributions**

650 M.C., Y.Z., J.H., and K.N. designed the study. M.C., Y.Z., Y.C., Y. Zha, S.L., G.C.,
 651 and M.L. collected samples and conducted experiments. M.C., Y.Z. conducted data
 652 analysis and produced the figures and tables. M.C., Y.Z., J.H., and K.N. wrote the
 653 manuscript. All authors revised the manuscript. M.C., Y.Z., J.H., and K.N. supervised
 654 the study.

655 Ethics statement

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

658 Competing interests

659 The authors declare that they have no competing interests

660 References

- 661 1. Almutairi, K., Nossent, J., Preen, D., Keen, H. & Inderjeeth, C. The global prevalence of
662 rheumatoid arthritis: a meta-analysis based on a systematic review. *Rheumatol. Int.* **41**,
663 863–877 (2021).
- 664 2. Barhum, L. What Does Rheumatoid Arthritis Progression Look Like?
665 <https://www.verywellhealth.com/rheumatoid-arthritis-stages-of-progression-4768891> (2021).
- 666 3. Smolen, J.S., Aletaha, D. & McInnes, I.B. Rheumatoid arthritis. *Lancet* **388**, 2023–2038
667 (2016).
- 668 4. Wells, P.M., *et al.* Associations between gut microbiota and genetic risk for rheumatoid
669 arthritis in the absence of disease: a cross-sectional study. *Lancet Rheumatol* **2**, e418–e427
670 (2020).
- 671 5. Zaiss, M.M., Joyce Wu, H.J., Mauro, D., Schett, G. & Ciccina, F. The gut-joint axis in
672 rheumatoid arthritis. *Nat. Rev. Rheumatol.* **17**, 224–237 (2021).
- 673 6. Jubair, W.K., *et al.* Modulation of Inflammatory Arthritis in Mice by Gut Microbiota Through
674 Mucosal Inflammation and Autoantibody Generation. *Arthritis Rheumatol* **70**, 1220–1233
675 (2018).
- 676 7. Takahashi, D., *et al.* Microbiota-derived butyrate limits the autoimmune response by
677 promoting the differentiation of follicular regulatory T cells. *EBioMedicine* **58**, 102913
678 (2020).
- 679 8. Smith, P.M., *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic Treg
680 cell homeostasis. *Science* **341**, 569–573 (2013).
- 681 9. Zhao, Y., *et al.* Hidden link in gut-joint axis: gut microbes promote rheumatoid arthritis at
682 early stage by enhancing ascorbate degradation. *Gut* (2021).
- 683 10. Pianta, A., *et al.* Evidence of the Immune Relevance of *Prevotella copri*, a Gut Microbe, in
684 Patients With Rheumatoid Arthritis. *Arthritis Rheumatol* **69**, 964–975 (2017).
- 685 11. Chen, J., *et al.* An expansion of rare lineage intestinal microbes characterizes rheumatoid
686 arthritis. *Genome Med.* **8**, 43 (2016).
- 687 12. Round, J.L. & Mazmanian, S.K. Inducible Foxp3⁺ regulatory T-cell development by a
688 commensal bacterium of the intestinal microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **107**,
689 12204–12209 (2010).
- 690 13. Chun, E., *et al.* Metabolite-Sensing Receptor Ffar2 Regulates Colonic Group 3 Innate
691 Lymphoid Cells and Gut Immunity. *Immunity* **51**, 871–884.e876 (2019).
- 692 14. Maeda, Y. & Takeda, K. Role of Gut Microbiota in Rheumatoid Arthritis. *J Clin Med* **6**, 60

693 (2017).

694 15. Aletaha, D., *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of
695 Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis*
696 *Rheumatol* **62**, 2569–2581 (2010).

697 16. Pedersen, H.K., *et al.* A computational framework to integrate high-throughput 'omics'
698 datasets for the identification of potential mechanistic links. *Nat Protoc* **13**, 2781–2800 (2018).

699 17. Li, D., *et al.* MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced
700 methodologies and community practices. *Methods* **102**, 3–11 (2016).

701 18. Truong, D.T., *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat.*
702 *Methods* **12**, 902–903 (2015).

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

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

707 21. Silvestre-Rangil, J., Bagán, L., Silvestre, F.J. & Bagán, J.V. Oral manifestations of rheumatoid
708 arthritis. A cross-sectional study of 73 patients. *Clin Oral Investig* **20**, 2575–2580 (2016).

709 22. Marriott, D., Stark, D. & Harkness, J. Veillonella parvula discitis and secondary bacteremia: a
710 rare infection complicating endoscopy and colonoscopy? *J. Clin. Microbiol.* **45**, 672–674
711 (2007).

712 23. Balakrishnan, B., Luckey, D. & Taneja, V. Autoimmunity-Associated Gut Commensals
713 Modulate Gut Permeability and Immunity in Humanized Mice. *Mil. Med.* **184**, 529–536
714 (2019).

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

717 25. Severmann, A.C., *et al.* An altered heparan sulfate structure in the articular cartilage protects
718 against osteoarthritis. *Osteoarthr Cartil* **28**, 977–987 (2020).

719 26. Henrotin, Y., Mathy, M., Sanchez, C. & Lambert, C. Chondroitin sulfate in the treatment of
720 osteoarthritis: from in vitro studies to clinical recommendations. *Ther. Adv. Musculoskelet. Dis.*
721 **2**, 335–348 (2010).

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

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

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

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

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

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

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

736 34. Manfredo Vieira, S., *et al.* Translocation of a gut pathobiont drives autoimmunity in mice and

737 humans. *Science* **359**, 1156–1161 (2018).

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

740 36. Zhao, Y., *et al.* Detection and characterization of bacterial nucleic acids in culture-negative
741 synovial tissue and fluid samples from rheumatoid arthritis or osteoarthritis patients. *Sci. Rep.*
742 **8**, 14305 (2018).

743 37. Siala, M., *et al.* Analysis of bacterial DNA in synovial tissue of Tunisian patients with reactive
744 and undifferentiated arthritis by broad-range PCR, cloning and sequencing. *Arthritis Res. Ther.*
745 **10**, R40 (2008).

746 38. Cox, C.J., Kempell, K.E. & Gaston, J.S. Investigation of infectious agents associated with
747 arthritis by reverse transcription PCR of bacterial rRNA. *Arthritis Res. Ther.* **5**, R1–8 (2003).

748 39. Scher, J.U., *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced
749 susceptibility to arthritis. *Elife* **2**, e01202 (2013).

750 40. Pucino, V., *et al.* Lactate Buildup at the Site of Chronic Inflammation Promotes Disease by
751 Inducing CD4(+) T Cell Metabolic Rewiring. *Cell Metab.* **30**, 1055–1074 (2019).

752 41. Rogatzki, M.J., Ferguson, B.S., Goodwin, M.L. & Gladden, L.B. Lactate is always the end
753 product of glycolysis. *Front. Neurosci.* **9**, 22 (2015).

754 42. Brown, N.A., Holt, D. & Webb, M. The teratogenicity of methoxyacetic acid in the rat.
755 *Toxicol. Lett.* **22**, 93–100 (1984).

756 43. Miller, R.R., Carreon, R.E., Young, J.T. & McKenna, M.J. Toxicity of methoxyacetic acid in
757 rats. *Fundam Appl Toxicol* **2**, 158–160 (1982).

758 44. Sparks, N.R. *The Embryotoxic Effects of Harm Reduction Tobacco Products on Osteoblasts*
759 *Developing from Human Embryonic Stem Cells*, (University of California, Riverside, 2018).

760 45. Li, P., *et al.* Metabolic Alterations in Older Women With Low Bone Mineral Density
761 Supplemented With *Lactobacillus reuteri*. *JBMR Plus* **5**, e10478 (2021).

762 46. Picchianti-Diamanti, A., *et al.* Analysis of Gut Microbiota in Rheumatoid Arthritis Patients:
763 Disease-Related Dysbiosis and Modifications Induced by Etanercept. *Int. J. Mol. Sci.* **19**, 2938
764 (2018).

765 47. Schirmer, M., *et al.* Linking the Human Gut Microbiome to Inflammatory Cytokine
766 Production Capacity. *Cell* **167**, 1125–1136.e1128 (2016).

767 48. Kverka, M., *et al.* Oral administration of *Parabacteroides distasonis* antigens attenuates
768 experimental murine colitis through modulation of immunity and microbiota composition.
769 *Clin Exp Immunol* **163**, 250–259 (2011).

770 49. Koh, G.Y., Kane, A.V., Wu, X. & Crott, J.W. *Parabacteroides distasonis* attenuates
771 tumorigenesis, modulates inflammatory markers and promotes intestinal barrier integrity in
772 azoxymethane-treated A/J mice. *Carcinogenesis* **41**, 909–917 (2020).

773 50. Hyatt, D., *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
774 identification. *BMC Bioinform* **11**, 119 (2010).

775 51. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein
776 or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

777 52. Nielsen, H.B., *et al.* Identification and assembly of genomes and genetic elements in complex
778 metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).

779 53. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**,
780 357–359 (2012).

- 781 54. Kanehisa, M., *et al.* Data, information, knowledge and principle: back to metabolism in
782 KEGG. *Nucleic Acids Res.* **42**, D199–205 (2014).
- 783 55. Aramaki, T., *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and
784 adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
- 785 56. Pedersen, H.K., *et al.* Human gut microbes impact host serum metabolome and insulin
786 sensitivity. *Nature* **535**, 376–381 (2016).
- 787 57. Song, G., *et al.* CD109 regulates the inflammatory response and is required for the
788 pathogenesis of rheumatoid arthritis. *Ann. Rheum. Dis.* **78**, 1632–1641 (2019).
- 789 58. Magoč, T. & Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome
790 assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
- 791 59. Caporaso, J.G., *et al.* QIIME allows analysis of high-throughput community sequencing data.
792 *Nat. Methods* **7**, 335–336 (2010).
- 793 60. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**,
794 2460–2461 (2010).
- 795 61. Cole, J.R., *et al.* The Ribosomal Database Project: improved alignments and new tools for
796 rRNA analysis. *Nucleic Acids Res.* **37**, D141–145 (2009).
- 797

798 Figure legends

799 **Fig. 1 | The analysis pipeline for detecting stage-specific gut-arthritis axis.** Gut
800 microbiome whole-genome sequencing (WGS) data and serum metabolomic data
801 from 27 healthy controls, 76 rheumatoid arthritis (RA), and 19 osteoarthritis (OA)
802 patients were used in the analysis. RA samples were divided into four stages including
803 RAS1 (n = 15), RAS2 (n = 21), RAS3 (n = 18), RAS4 (n = 22) according to the
804 rheumatoid diagnostic score. **a**, Microbial genes, predicted from WGS sequences,
805 were clustered into metagenomic species (MGS) for taxonomic annotation and
806 mapped to KEGG orthology for functional annotation. Ten KEGG modules that were
807 significantly correlated with arthritis (Healthy = 0, OA = 1, RA = 2, $P < 0.05$, FDR <
808 0.05, Mann-Whitney-Wilcoxon test) were remained to describe the stage-specific
809 functional pattern. Additionally, the leave-one-out analysis was used to determine the
810 MGS that most drove the correlation of these ten KEGG modules with arthritis
811 (Methods). KEGG maps of these specific modules, their KOs' abundances, and the
812 corresponding driving MGS abundances were plotted to show the stage-specific
813 functional driving genes variation. **b**, A total of 232 classified microbial species,
814 identified from WGS sequences, were used to describe the stage-specific taxonomic

815 pattern. **c**, In total, 277 serum metabolites were used to describe the stage-specific
 816 metabolomic pattern. **d**, Random forest algorithm was performed on 6224 KOs, 232
 817 microbial species, and 277 serum metabolites from (a–c) to detect the stage-specific
 818 biomarkers at the multi-omics level and the AUROC in classifying stage-specific
 819 arthritis samples from healthy controls.

820 **Fig. 2 | Distinct stage-specific taxonomic patterns in arthritis progression. a,**

821 Linear discriminant analysis to separate the samples of healthy, OA, and RA groups.

822 The length and direction of the arrows indicate the normalized scalings for each of the

823 microbial species. **b**, Each row represents a species with significantly different

824 abundances ($P < 0.05$, FDR < 0.05) between the healthy and RA groups or between

825 the healthy and OA groups. Each column represents a sample from one of the groups

826 including the healthy ($n = 27$), RAS1 ($n = 15$), RAS2 ($n = 21$), RAS3 ($n = 18$), RAS4

827 ($n = 22$), and OA ($n = 19$) groups. The color of each heatmap unit represents the

828 scaled abundance of a certain species in a specific sample. Species are colored for

829 significant elevation (red) or depletion (green) in the arthritis groups, in comparison

830 with the healthy groups. **c**, Three apices of the triangle in the ternary diagram

831 represent components of healthy, OA, and RA groups. Microbial species in (b) are

832 plotted as dots inside the triangle, the size of which indicates the relative abundance.

833 For each dot, moving away from the apex, the proportion of the component at the

834 apex decreases as the proportion of the remaining two increases. **d**, In total, 232

835 classified species whose relative abundance was more than 1×10^{-6} are shown in a

836 phylogenetic tree, grouped in the phyla Fusobacteria, Proteobacteria, Actinobacteria,

837 Bacteroidetes, and Firmicutes. In the outer circles, species at each arthritis stage are

838 marked for significant elevation ($P < 0.05$, orange; FDR < 0.05 , red) or depletion ($P <$

839 0.05 , light green; FDR < 0.05 , dark green) in abundance, as compared to those of

840 healthy controls. The innermost circle shows species relative abundances averaged

841 over all samples. Box plots show the relative abundances of species recorded in (b)

842 and (c), and others that have been correlated with RA in previous work. The order and

843 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 $1.5 \times$ 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.

Fig. 3 | Distinct stage-specific functional patterns in arthritis progression. Gene abundances were assessed for elevation or depletion in each of the arthritis stages, RAS1 ($n = 15$), RAS2 ($n = 21$), RAS3 ($n = 18$), RAS4 ($n = 22$), and OA ($n = 19$) compared to the healthy controls ($n = 27$). **a**, Relative abundance of KO genes in the KEGG modules that were significantly correlated with arthritis ($P < 0.05$, FDR < 0.05). KO genes with a prevalence of 5% or higher are shown. 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 $P < 0.05$. **b**, KO genes involved in specific KEGG pathway modules in (a) are shown in the KEGG pathway maps. Each box in a pathway represents a KO gene and is marked in red for elevation or in blue for depletion at any of the stages compared to healthy controls. Bar plots show relative gene abundances averaged over samples within each of the five groups (healthy (H), RAS1 (S1), RAS2 (S2), RAS3 (S3), RAS4 (S4), and OA) and are colored according to the values. Each KO gene is composed of MGS genes represented by circles. The sizes and colors of the circles are proportional to the relative abundances of the MGS genes. MGS genes are grouped into one row and indicated by the taxonomic name. The three MGS that most drove the correlation of the KEGG modules with arthritis types are shown. 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

873 $P < 0.01$; –, depletion at $P < 0.05$. The boxes represent 25th–75th percentiles, black
874 lines indicate the median, and whiskers extend to the maximum and minimum values
875 within $1.5 \times$ the interquartile range. All the statistical significance was calculated by
876 the Mann-Whitney-Wilcoxon test.

877 **Fig. 4 | Stage-specific biomarkers for discriminating arthritis progression from**
878 **healthy status at a multi-omics level.** A random forest algorithm was performed on
879 6224 KOs, 232 microbial species, and 277 serum metabolites to detect the
880 stage-specific biomarkers in the arthritis progression including RAS1 (a), RAS2 (b),
881 RAS3 (c), RAS4 (d), and OA (e). In each of the panels, the Gini importance of the top
882 five most discriminant biomarkers of microbial species, KOs, or metabolites are
883 displayed. Boxes represent the interquartile range between the first and third quartiles
884 and the line inside represents the median. Whiskers denote the lowest and highest
885 values within the $1.5 \times$ interquartile range from the first and third quartiles,
886 respectively. Boxes are marked in a specific color to show the significant elevation (P
887 < 0.05 , red, Mann-Whitney-Wilcoxon test) or depletion ($P < 0.05$, blue,
888 Mann-Whitney-Wilcoxon test) of the biomarkers in each of the arthritis stages
889 compared to the healthy group. The ROC curves of the random-forest model using
890 microbial species, KOs, or metabolites were plotted, with AUC calculated by 10
891 randomized 10-fold cross-validation. The color of the curve represents the categories
892 of the used features.

893 **Fig. 5 | Stage-specific gut microbial species, serum cytokines, and serum**
894 **metabolites alteration in gut-joint axis. a,** The alteration of gut microbiome
895 taxonomic and functional composition would interact with human joint through serum
896 metabolites perturbation. The most dominant elevation and depletion of microbial
897 functionality in arthritis, or microbial species or serum metabolites at each arthritis
898 stage are shown, as compared to those in the healthy group. The perturbed
899 metabolites and elevated serum cytokines such as IL-6 and TNF- α were transited to
900 human joints to promote inflammation and bone erosion. Different arthritis stages had
901 inflammation or bone erosion of different severity and corresponding different

902 microbial patterns. **b**, The dot plots show stage-specific abundance or concentration
 903 (mean $\pm$ s.e.m.) of microbial species, serum cytokines, and serum metabolites, which
 904 are specified in (a). Four RA stages are connected to display the variance. For
 905 microbial species and serum metabolites, dots are colored differently if the features
 906 are significantly elevated (red) or significantly depleted (blue), as compared to those
 907 of the healthy group. For serum cytokines, dots of RA stages are colored in red for
 908 significant elevation compared to the OA group. ***, $P < 0.005$; **, $P < 0.01$; *, $P <$
 909 0.05, Mann-Whitney-Wilcoxon test.

910 **Fig. 6 | The RA-stage-specific changes were related to RA pathogenesis.** **a**, Certain
 911 potential mechanisms by which gut microbial dysbiosis is involved in RA
 912 pathogenesis are displayed. Gut microbial variation drove the changes in microbial
 913 functionality that influenced the metabolites, which might promote RA progression
 914 via hurting bone tissue and increasing inflammation. The driving species, functions,
 915 and metabolites were extracted from **Fig. 3b**. The red or blue box of metabolites
 916 represents their speculated elevation or depletion according to the functional variation.
 917 The dotted line represents the speculated effects of microbial and metabolic variation
 918 on arthritis pathogenesis, with a bracket including the literature. **b**, The stage-specific
 919 changes at each of the RA stages that would promote RA progression. The major
 920 contributions of each RA stage, reflected by multi-omics, are displayed. The grey box
 921 showed that although these processes were the most notable changes at each of the
 922 RA stages, they kept ongoing throughout the RA progression. (a) characterized a few
 923 links between microbial dysbiosis and the joint of (b) throughout the RA progression.

Recruited cohort

Healthy (27)
RA (76)
OA (19)

S1 (15)
S2 (21)
S3 (18)
S4 (22)

Gut-joint axis

WGS sequences

MEGAHIT, Prodigal

MetaPhlAn2

a

Microbial genes (4M)

Taxonomy

Functionality

MGS (553)

KEGG orthologs (6224)

KEGG modules (404)

Leave-one-out analysis

MGS that most drive the correlation of
KEGG modules with arthritis

(12)

Arthritis association
filtering (FDR < 0.1)

Stage-specific
functional pattern

Stage-specific
functional driving genes variation

b

Microbial species (232)

Stage-specific
taxonomic pattern

c

Serum metabolites (277)

Stage-specific
metabolomic pattern

d

Random forest
Stage-specific
biomarkers

a **b**

a**RAS1** **b****RAS2** **c****RAS3** **d****RAS4** **e****OA**

Significantly elevation
compared to the healthy group ($P < 0.05$)

Significantly depletion
compared to the healthy group ($P < 0.05$)

Not significant

a

b

a **b**