Alterations in the gut microbiome in inflammatory arthritis implicate key taxa and metabolic pathways across arthritis phenotypes

One sentence summary: Gut microbiome configurations and activity in several types of inflammatory arthritis were found to be similar to those in gastrointestinal disease, with additional strain-specific links to iron and vitamin metabolism

Authors: Kelsey N Thompson[1,2,3], Kevin S Bonham[1,2], Nicholas E Ilott[4], Graham J Britton[5], Paula Colmenero[4], Samuel J Bullers[4], Lauren J McIver[1,3], Siyuan Ma[1,2], Long H Nguyen[1,3,6], Andrew Filer[7,8], India Brough[4,9], Claire Pearson[4], Caroline Moussa[4], Vinod Kumar[4], Lilian H Lam[4], , Matthew A Jackson[4], April Pawluk[1,3], The Inflammatory Arthritis Microbiome Consortium (IAMC) investigators group, Serafim Kiriakidis[9], Peter C Taylor[9], Lucy R Wedderburn[10,11,12], Brian Marsden[4], Stephen P Young[13], Dan R Littman[14], Jeremiah J Faith[5], Arthur G Pratt[15,16,17], Paul Bowness[9] , Karim Raza[7,8,13], Fiona Powrie[4], and Curtis Huttenhower[1,2,3,18]

Affiliations:

[1]Department of Biostatistics, T.H. Chan School of Public Health, Harvard University; Boston, MA, 02115, USA
[2]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard; Cambridge, MA, 02142, USA
[3]Harvard Chan Microbiome in Public Health Center, Harvard T. H. Chan School of Public Health; Boston, MA, 02115, USA
[4]Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedic, Rheumatology and Musculoskeletal Sciences, Oxford University; Oxford, OX3 7FY, UK
[5]Marc and Jennifer Lipschultz Precision Immunology Institute and Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA
[6]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School; Boston, MA, 02114, USA
[7]Institute of Inflammation and Ageing, College of Medical and Dental Sciences, University of Birmingham, Queen Elizabeth Hospital; Birmingham, B15 2TT, UK
[8]MRC Versus Arthritis Centre for Musculoskeletal Ageing Research and Research Into Inflammatory Arthritis Centre Versus Arthritis, University of Birmingham; Birmingham, UK
[9]Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford; Oxford, OX3 7LD, UK
[10]Centre for Adolescent Rheumatology Versus Arthritis, University College London, UCLH, and GOSH; London, UK
[11]NIHR Great Ormond Street Biomedical Research Centre, University College London; London WC1N 1EH, UK
[12]UCL GOS Institute of Child Health, University College London; London WC1N 1EH, UK
[13]Department of Rheumatology, Sandwell & West Birmingham NHS Trust; West Bromwich B71 4HJ, UK
[14]Howard Hughes Medical Institute and the Kimmel Center for Biology and Medicine of the Skirball Institute, New York University School of Medicine; New York, NY 10016, USA
[15]Translational and Clinical Research Institute, Newcastle University; Newcastle upon Tyne NE2 4HH, UK
[16]Research into Inflammatory Arthritis Centre Versus Arthritis, Newcastle Birmingham; Glasgow, and Oxford, UK
[17]Department of Rheumatology, Musculoskeletal Services Directorate, Newcastle upon Tyne Hospitals NHS Trust; NE7 7DN, UK
[18]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health; Boston, MA, 02115, USA

Abstract

Musculoskeletal diseases affect up to 20% of adults worldwide. The gut microbiome has been implicated in inflammatory conditions, but large-scale metagenomic evaluations have not yet traced the routes by which immunity in the gut affects inflammatory arthritis. To characterize the community structure and associated functional processes driving gut microbial involvement in arthritis, the Inflammatory Arthritis Microbiome Consortium investigated 440 stool shotgun metagenomes comprising 221 adults diagnosed with rheumatoid arthritis, ankylosing spondylitis, or psoriatic arthritis, and 219 healthy controls and individuals with joint pain without an underlying inflammatory cause. Diagnosis explained ~2% of gut taxonomic variability, which is comparable in magnitude to inflammatory bowel disease. We identified several candidate microbes with differential carriage patterns in patients with elevated blood markers for inflammation. Our results confirm and extend previous findings of increased carriage of typically oral or inflammatory taxa, and decreased abundance and prevalence of several typical gut clades, indicating that distal

14 inflammatory conditions correspond in similar alterations to the gut microbial composition as do
15 local conditions. We identified several differentially encoded pathways in the gut microbiome of
16 patients with inflammatory arthritis, including changes in vitamin B salvage and biosynthesis and
17 and enrichment of iron sequestration. While several of these changes characteristic of
18 inflammation could have causal roles, we hypothesize that they are mainly "positive feedback"
19 responses to changes in host physiology and immune homeostasis. By taxonomic to functional
20 alterations, this work thus expands our understanding of the key shifts in the gut ecosystem in
21 response to systemic inflammation during arthritis.

Introduction

23 Alterations to the gut microbiome have been implicated in several inflammatory diseases,
24 particularly in the gastrointestinal tract (GIT), including the inflammatory bowel diseases (IBD)
25 and colorectal cancer (CRC)(*1, 2*). Although the role of gut microbes in other inflammatory
26 conditions such as type 1 diabetes (T1D) and metabolic syndrome have recently come under
27 investigation(*3, 4*), their influences on or responses to systemic inflammation or disease
28 progression remain poorly elucidated. Inflammatory musculoskeletal arthropathies stand to
29 benefit from a better understanding of gut microbial ecology, both as an early biomarker for
30 diagnosis of these conditions, and as a potential new route for therapy. Arthropathies including
31 rheumatoid arthritis (RA), ankylosing spondylitis (AS), and psoriatic arthritis (PsA) affect over 50
32 million adults worldwide(*5-7*), who currently have no curative treatment options. Thus,
33 understanding their corresponding alterations within the gut microbiome is essential to both the
34 underlying basic biology driving systemic inflammation and clinical routes of arthritis treatment.

35 The etiology of many of the subtypes of arthritis can be traced back to aberrant immune
36 responses, which may be triggered or sustained by acute or long-term interactions with gut
37 microbial populations(*8*). This is true over and above human genetic contributions, which include
38 variants of the human leukocyte antigen (HLA) family(*9*). In RA, heritability is estimated at 60%,
39 although without clearly-resolved causal loci(*10-13*). Conversely, the heritability of AS
40 approached 90%, with HLA-B27 carriage the strongest genetic risk factor(*14-16*). Additionally,
41 smoking has been established as a likely trigger of RA, representing at least one specific
42 interaction between environmental and genetic factors in arthritis etiology(*13, 17*). The
43 microbiome is one of the most proximal forms of "environment", and indeed many arthritis risk
44 alleles such as HLA are (like those of IBD) known to be involved in microbial interactions or
45 immune sensing(*18*).

46 Since arthritis pathology is localized in the periphery, all of these arthropathies represent cases
47 in which any involvement of the gut microbiome would be "transmitted" systemically through
48 biochemical and immune-mediated signals. Research on this so-called "gut-joint-axis" dates back
49 to the 1890s, when researchers hypothesized that arthritic conditions could be caused by
50 *Mycobacterium* infections(*19*). Murine models have furthered this hypothesis by showing that
51 microbial disease triggers are required for SpA type arthritis to develop(*20-22*) and that gut
52 microbial colonization is necessary for Th17 differentiation (protecting germ-free mice from
53 disease)(*23*). Several strong indicators of the "gut-joint-axis" exist in humans as well, including
54 subsets of patients with chronic IBD exhibiting increased risk of peripheral arthritis(*24, 25*),
55 reactive arthritis occurring after pathogen infections(*26*), and the induction of autoreactive
56 cartilage degradation by specific bacterial strains(*25*). Several studies in smaller human
57 populations, primarily studied via 16S rRNA gene amplicon (16S) sequencing, found
58 compositional alterations of the gut microbiome in patients with RA, PsA, and AS(*27-41*). These
59 included the presence of clades that are frequently pathogenic, increased abundance of typically
60 oral microbes in the gut, and altered abundance of typical human gut clades(*39-42*). However,

61 there is no substantial agreement on which dysbioses are hallmarks of systemic inflammation in
62 arthritis. Additionally, 16S-based profiles do not provide direct insight into the functional
63 implications of microbial compositional changes, and thus far the agreement in functional changes
64 from the few shotgun studies is limited but have identified sweeping changes(*34, 36, 39-41, 43*).
65 Therefore, a comprehensive understanding of the role of the gut microbiome in arthritis
66 development and persistence is still lacking, which has the potential to better-support early
67 disease detection, prevention, or later-stage therapy.

68 Here, we introduce the work of the Inflammatory Arthritis Microbiome Consortium (IAMC), which
69 includes analysis of shotgun metagenome profiles spanning 440 subjects with RA, AS, PsA, and
70 controls without inflammatory arthritis. We assessed the taxonomic and functional landscape of
71 the resulting gut microbiomes to elucidate key ecological and biochemical shifts linked to host
72 inflammatory responses and clinical arthritis phenotypes. In patients with inflammatory arthritis,
73 the overall compositional and functional profiles of the gut microbiome were significantly altered.
74 We identified enrichment of typically oral, pro-inflammatory, and mucin-degrading microbes, with
75 a corresponding decrease in several typical human gut-resident clades. Notably, several strains
76 of *Ruminococcus gnavus* isolated from human patients induced more severe phenotypes when
77 inoculated into mice. Further, several alterations in microbial community function were identified,
78 including the differential encoding of vitamin B salvage and biosynthesis and the encoding of folic
79 acid metabolism pathways. Similar to other local and distal inflammatory diseases, iron
80 scavenging was enriched in patients with current inflammation across heme, non-heme, and
81 siderophore-based mechanisms. While these findings point to pathways and molecules of interest
82 and will serve as an important resource for hypothesis generation, future work will be required to
83 determine if these consistent functional changes occur causally, in response to inflammatory
84 arthritis, or both. At the least, our findings of community level taxonomic and functional alterations
85 in the gut microbiome implicate an interplay between host genetics, immune system, and gut
86 microbiome over the course of initiation, progression, and severity of arthritis.

87 Results

88 We recruited 440 adults (ages 20-93) from different clinical locations in the United Kingdom,
89 Oxford (primarily AS patients), Birmingham (primarily RA patients), and Newcastle (primarily RA
90 patients), who met classification criteria for one of three arthritis subtypes or were included in the
91 non-inflammatory joint pain control group (**Fig. 1A-B**). Patient diagnoses included primarily
92 treatment naïve rheumatoid arthritis (RA, n=119), axial spondyloarthritis/ankylosing spondylitis
93 (AS, n=67), psoriatic arthritis (PsA, n=35), and non-inflammatory joint pain (typically fibromyalgia
94 [NIJP], n=54), as well as age-matched healthy controls (HC, n=165) (**Fig. 1A, Supplemental
95 Table S1**). Whilst some members of the NIJP category had low-titer autoantibodies for either RF
96 or anti-CCP (never both), none were considered by consulting rheumatologists to have clinically
97 suspect arthralgia with respect to RA(*44*).

98 Earlier studies have identified alterations in gut microbial taxonomic profiles with both arthritis
99 diagnosis and local and distal inflammation(*27-33*). To expand these results, we focused on
100 patient diagnosis and current level of disease activity as primary outcomes while adjusting for
101 relevant clinical covariates including patient age, current arthritis-related drug use, and disease
102 duration, as well as technical confounders such as sequencing batch and clinical site
103 (**Supplemental Fig. S2A-I**). Only a small subset of patients (58/440) reported use of antibiotics
104 in the last six months, ~half (22) within the previous two months and none at the time of sampling.
105 These were equally spread across all patients and HC's (**Table 1**); such non-recent antibiotics
106 use corresponded with very little overall variation (PERMANOVA R^2=0.0049), and no individual
107 features were significantly associated with antibiotics use (MaAsLin). Antibiotic covariates were

108 thus omitted from further analyses. We defined disease activity using two variables: (1) discretized
109 C-reactive protein values (CRP) as a marker of current systemic inflammation (**Fig. 1D**), and (2)
110 serum hemoglobin levels, as many patients with inflammatory arthritis also experience anemia as
111 a feature of chronic disease (**Supplemental Fig. S2I**).

112 Individuals' disease activities varied from low to high disease activity levels [Bath Ankylosing
113 Spondylitis Disease Activity Index (BASDAI) = 0.08-8.4 (AS-specific measure), Disease Activity
114 Score 28 for RA with CRP (DAS28-CRP) = 1.54-8.01 (RA/PsA-specific measure)] (**Fig. 1C**,
115 **Supplementary Table S1, Supplemental Fig. S2C-D**). While not ideal measures of
116 inflammation, CRP levels for each subject provide a measure that was collected consistently
117 across all population cohorts and ranged from 0-167 mg/L (**Supplemental Fig. S2C-D**). Thus, as
118 the most consistently collected measure, we used CRP as a proxy for systemic inflammation;
119 when available, it compared favorably to more direct measures such as BASDAI for AS or DAS28
120 for RA. Due to the asymmetric nature of these data, we categorized these patients using tertiles
121 into three categories: *not inflamed* (0-4 mg/L), *some inflammation* (4-10 mg/L), and *inflammation*
122 (>10 mg/L). Healthy controls only contributed fecal samples, not blood, and we could thus not
123 quantify CRP levels from these individuals and categorized all controls into *not inflamed* (**Fig 1D**).
124 From here on, we refer to the discretized CRP-value for systemic inflammation as simply
125 "inflammation." Anemia was also quantified in this population by current hemoglobin levels, with
126 anemia called when hemoglobin was less than 120g/L or 135g/L for females and males,
127 respectively (**Supplemental Fig. S2I**). Human leukocyte antigen B27 (HLA-B27) status was
128 quantified as either negative or positive, but only for the patients from Oxford (AS patients, n=67).
129 In addition, rheumatoid factor (RF) and anti-cyclic citrullinated peptide (anti-CCP) status was
130 categorized as negative or positive for RA patients (n=113) **(Supplemental Table 1;**
131 **Supplemental Fig. S2F-H**).

132 **The human gut microbiome is altered in inflammatory arthritis**
133 Significant alterations in the overall composition of the gut microbiome were identified in patients
134 with inflammatory arthritis. Patient diagnosis explained a maximum of 1.6% and 2.3% of the
135 compositional differences in the taxonomic and functional microbial profiles, respectively, after
136 adjusting for the sequencing batch (Bray-Curtis PERMANOVA; FDR q-value = 0.003 and 0.006).
137 In pairwise comparisons, these results were driven largely by differences in the RA patients
138 (**Supplemental Fig. S3**). Categorized CRP-values, which represent the current amount of
139 inflammation a patient is experiencing, accounted for a maximum of 1.2% (FDR q-value = 0.003)
140 and 2.0% (FDR q-value = 0.006) of the variation in the composition of the taxonomic and
141 functional profiles, respectively (**Fig. 1E; Supplemental Fig. S3**). Inflammation thus explained a
142 small but notable shift in the overall gut microbial composition, not greatly below the amount often
143 observed in IBD(*1*) (**Fig. 1F**). Clinical measures of inflammation such as the patients' DAS28-
144 CRP and BASDAI also explained similar amounts of variation within the gut ecology (**Fig. 1E**).
145 This indicates that systemic inflammation during arthritis, as characterized by either disease-
146 specific markers or circulating measures in all patients, corresponds with a significant amount of
147 variation in the patients' gut microbiomes. Intriguingly, similar amounts of variation were also
148 explained by a patient's hemoglobin concentrations (g/L) (Taxonomy; R^2 = 1.1% and FDR q-value
149 = 0.003, Pathway; R^2 = 1.2% and FDR q-value = 0.009, **Fig. 1E**). Similar magnitudes of
150 significance also demonstrate a consistent, but diverse, coupling of taxonomic and functional
151 aspects of the gut microbiome, as expected. HLA-B27, anti-CCP and RF-status all did not induce
152 significant alterations in the overall composition of the gut microbiome (Bray-Curtis PERMANOVA
153 taxonomy; FDR q-value >0.01). Taken together, these results indicate that patients with
154 inflammatory arthritis do harbor broadly different configurations of microbes within their gut when
155 compared to similarly-aged healthy controls, consistent with previous studies(*28, 34, 42*).

Microbial taxonomic alterations in rheumatoid arthritis and ankylosing spondylitis

We identified several taxa significantly associated with inflammation, diagnosis, and anemia or more disease-specific markers of inflammation (e.g. BASDAI or DAS28-CRP) that paralleled changes previously observed in dysbiotic individuals with IBD[1], including the clades *Streptococcus sp.*, *Escherichia coli*, and *Ruminococcus gnavus*[45] (**Fig. 2A; Supplemental Fig. S4-8**). Examining the prevalence of these organisms across patients, it appears that *E. coli* and *R. gnavus* may exhibit a high abundance but low prevalence phenotype, in which a small number of patients had substantially higher abundances of these taxa. Previously, this pattern was observed with *Prevotella copri* in treatment-naïve RA patients[29, 34, 46], but that was not the case in this cohort (**Supplemental Fig. S9** and **Supplemental discussion**). Several of the clades increased during inflammation are more commonly identified in the oral cavity[47-49] than in the gut, including *Streptococcus mutans*, *S. vestibularis*, *S. salivarius*, and *Bifidobacterium dentium* (**Fig. 2B-C; Supplemental Fig. S5-6, 8**). For these tests, we were careful to adjust for proton pump inhibitor (PPI) usage, which has been hypothesized to facilitate the transversion of oral taxa into the gut[50, 51]. However, only 11 patients out of 275 were documented to be actively taking PPIs at the time of sample collection, and we thus do not believe this to be the mechanism. We do not have information on the oral health status (e.g. periodontal disease) of these patients, previous studies have found that the patients with RA were four times as likely to have poor dental health[52], which strengthens the association of these different types of microbial disruption, but leaves their respective causalities unclear.

Patients with IBD and T1D have both showed reduced abundance of the species *Faecalibacterium prausnitzii, Roseburia intestinalis, Eubacterium rectale* and *Alistipes putredinis*[1, 53-55], and we observed a similar decrease in our study for AS and RA patients, either by our proxy marker for inflammation or by disease phenotype or disease-specific markers (**Fig. 2C-D; Supplemental Fig. S4-5, 8**)[55-61]. In particular, *F. prausnitzii* and *R. intestinalis* lost both abundance and prevalence in patients with current inflammation, while *E. rectale* abundance was observed to be tightly coupled with hemoglobin levels. These microbes are generally considered to be both highly responsive to inflammation and themselves anti-inflammatory by routes such as SCFA production[62-64], making the causality of these changes difficult to untangle observationally. However, it is striking that gut microbial changes observed here for systemic inflammation during arthritis were both taxonomically and functionally similar to those occurring during gastrointestinal inflammation which has been hypothesized to occur due to changes in oxygen availability in the gut ecosystem[65, 66].

R. gnavus has been implicated in many inflammatory conditions[67-70] and has been researched more extensively in IBD[71-73]. Additionally, using both single nucleotide variants (SNVs, via StrainPhlAn[74]) and differences in pangenome-wide gene content (via PanPhlAn[75]), we identified phylogenetic structures that were significantly enriched in AS and RA patients (denoted *Clade One;* Kimura 2-parameter distance, PERMANOVA; $R^2 = 0.18$, FDR q-value = 0.01) when compared the NIJP and HC individuals (which tended to carry members of *Clade Other*; **Fig. 3B** and **Supplemental Fig. S12-14**). To strengthen these results, an isolate from one RA patient also fell into *Clade One*, along with several isolates from a previously published IBD cohort[71]. These results indicate that the presence of inflammation both locally and at distal locations in the host can correlate with structural, and potentially functional, changes in the gut microbiome. Other species tested did not exhibit the same subclade structuring as *R. gnavus* in this population (**Supplemental Fig. S10**).

Notably, isolates from *Clade One* specifically enhanced inflammatory phenotypes when introduced into a mouse model of arthritis. New isolates (**Fig. 3B**) were derived from participant fecal samples and inoculated into previously germ-free SKG mice. Two to three weeks after the

204 introduction of these monocultures of *R. gnavus*, curdlan was injected to induce arthritis
205 symptomology. The presence of isolates from *Clade One* in the gut were able to significantly
206 potentiate the severity of arthritis-like symptoms in the SKG mouse, using both joint diameter and
207 disease score as indicators of severity (**Fig. 3C-E**). This showed that the presence of these strains
208 of *R. gnavus* is sufficient to induce a more severe phenotype, supporting its likely interaction with
209 the immune system, as previously postulated(*72, 76*), and furthering arthritis symptoms.

210 Functional profiling reveals consistent functional alterations across all subtypes of
211 arthritis

212 We observed increased carriage of folate metabolism pathway and enzymes in arthritis patients
213 and those individuals with current high levels of systemic inflammation (e.g. PWY-3841: Folate
214 transformation II; linear model not inflamed vs. inflammation; coef = 0.4, q-value = 0.15; **Fig. 4A-**
215 **C; Supplemental Fig. S15-16**). However, carriage of the methionine cycle (which typically
216 includes production of putrescine and homocysteine) was significantly decreased across both RA
217 and inflammation (e.g. PWY-6151: S-adenosyl-L-methionine cycle I; linear model not inflamed vs.
218 inflammation; coef = -0.03, FDR q-value = 0.17) (**Fig. 4A-C**). Alterations to nucleotide and amino
219 acid pathways downstream of these processes are enumerated in **Supplementary Fig. S19**.
220 These lines of evidence suggest a dysregulation of folate metabolism in the gut ecosystem during
221 arthritis, but interestingly this was distributed among a variety of different potential encoding
222 organisms in different subjects (**Supplemental Fig. S16**).

223 Patient hemoglobin levels and inflammation status were also associated with decreases in gut
224 microbial carriage of vitamin B12 salvage and biosynthesis pathways (e.g EC 1.2.1.54: Precorrin-
225 6A reductase; linear model Normal vs. Anemia; coef = -0.17, FDR q-value = 0.14), in tandem with
226 these disruptions in vitamin B9 (folate) and its interconnected pathways (**Fig. 4A-B, D**). Vitamin
227 B12 is a required cofactor in the link between the folate and methionine cycles(*77*) (**Fig. 4B**).
228 These alterations in the encoding of B12 salvage appear to be due to mainly (but not entirely) to
229 shifts in the relative abundance of *Eubacterium rectale* as noted above (**Supplemental Fig. S17**).
230 Additionally, the metagenomic abundances of several other enzymes associated with vitamin B
231 metabolism were also significantly linked to patient diagnosis, inflammation status, or anemia
232 status, including vitamin B1 (thiamin/thiazole), vitamin B2 (flavin), vitamin B6 (pyridoxine), and
233 vitamin B7 (biotin) (**Supplemental Fig. S15, 18, 20**).

234 Alterations of other gut metagenomic pathways regulating metabolic cofactors were also identified
235 among inflammatory arthritis patients. Genes encoding several enzymes linked with microbial iron
236 sequestration were enriched in inflamed patients, as well as in RA and AS patients specifically
237 (e.g. ENTBACSYN-PWY: Enterobactin biosynthesis; linear model not inflamed vs. inflammation;
238 coef = 0.62, q-value = 0.00014). These included genes encoding ferrochelatase, bacterial non-
239 heme ferritin, ferroxidase, and heme biosynthesis (at the pathway level) (**Fig. 5A-B**). This agrees
240 with results in earlier, smaller RA cohort in which oral and gut capacity for iron transport was
241 disrupted(*34*). However, this was previously attributed specifically to Klebsiella spp.; in contrast,
242 as with disruptions in folate metabolism, we found contributions to iron sequestration to be
243 encoded by diverse taxa (**Fig. 5C** and **Supplemental Fig. S21**), and for non-heme mechanisms,
244 no one clade was individually associated with their differential carriage. This again indicates that
245 dysbiosis of these processes can be distributed among different microbes in different people,
246 and/or that there are strain-level differential carriage within taxa, such as with ABC transporters
247 within *R. gnavus* (**Fig. 3C**). However, heme related mechanisms appear to be driven mainly by
248 the high abundance phenotype described above for *E.coli* in this population, as it is a major
249 contributor to these functions (**Fig. 5C** and **Supplemental Fig. S21**).

250 Several other functional classes, including both pathways and enzymes, exhibited differential
251 metagenomic carriage either by inflammation status or patient diagnosis (**Supplemental Fig.**
252 **S22-26**). The gene classes most highly associated with both RA and inflammation were those
253 relating to the production of isoprenoids or volatile hydrocarbons (VOCs) (**Supplemental Fig.**
254 **S22**). Microbial gene families that explicitly interact with host immunity were also differentially
255 carried during arthritis, such as cell wall remodeling proteins and oxidative stress response
256 (**Supplemental Fig. S23-26**). Finally, enzymes involved in the oxidative stress response including
257 methanogenesis, glutathione, and peroxiredoxin (**Supplemental Fig. S23**) and of interest to the
258 current short-chain fatty acid literature(*78*), 3-hydroxybutyryl-CoA dehydrogenases and short-
259 chain acyl-CoA dehydrogenase enzymes were differentially carried in inflammation
260 (**Supplemental Fig. S24**). Overall, alterations to the functional landscape of the gut ecosystem
261 indicated disruptions in several key metabolism pathways during inflammatory arthritis.

Discussion

263 Here, we present the findings of a large cross-sectional study of adults with inflammatory arthritis
264 diagnosis (and control participants), investigating alterations in gut microbiome composition and
265 function associated with disease status and inflammation. Strikingly, the signals we detected
266 associated with this family of systemic inflammatory conditions largely paralleled those identified
267 in diseases defined by gastrointestinal inflammation, such as IBD. Changes in microbial taxa,
268 functions (pathways and individual gene families), and in some cases even strains (e.g.
269 *Ruminococcus gnavus*) were shared between arthritis patients and other inflammatory diseases
270 such as IBD, T1D, and other metabolic disorders. These changes were largely consistent among
271 RA, AS, and PsA, which was not captured by previous studies. In addition to the initial results
272 presented here, the corresponding large shotgun metagenomic and clinical dataset offers the
273 ability for further hypothesis generation and testing, including the potential for identification of
274 novel arthritis therapeutic targets.

275 In particular, several previously-suggested "pro-inflammatory" microbes were enriched here
276 during arthritis(*1*). This was particularly true for *E. coli*, which had an especially unique influence
277 on the corresponding community functional potential, and has been previously shown to be
278 enriched in many conditions including RA(*43, 79*). Mucin-degrading microbes such as *R.*
279 *gnavus*(*68, 69*) were also differentially carried and functional during disease, down to the
280 subspecies level (**Fig. 2**). Recently, a substantial number of studies have found direct
281 associations between *R. gnavus* and inflammation, including in arthritis(*84, 85*). Potential
282 mechanisms include direct interaction of *R. gnavus* with the host immune system through
283 extracellular proteins(*86*). Notably, a subset of phylogenetically distinct *R. gnavus* isolates from
284 an individual in this cohort with RA and isolates from a healthy control and two IBD individuals
285 were sufficient to increase arthritis severity when introduced into SKG, both supporting their
286 causality and agreeing with previous subclade results (**Fig. 3B-E**). Finally, the presence of
287 characteristically oral taxa in the gut microbiome of patients with chronic inflammation has been
288 well documented, including in patients with IBD, UC, CRC, and metabolic disorders(*1, 48, 87, 88*).
289 A few studies on patients with both AS and RA have also identified increases in streptococci in
290 the gut(*21, 42*), similar to what was observed in our population (**Fig. 2**).
291
292 These examples represent two ways in which our results generally agree with previous studies of
293 the gut microbiome in inflammatory arthritis(*41, 43, 79*). We also observed similar broad patterns,
294 such as the loss of typical gut consortia and increasing abundance of oral taxa and clades
295 associated with gastrointestinal inflammation(*34-41, 43*). While we were slightly hindered by
296 differences in collection targets within our disease subtypes (e.g. treatment-naive early RA

297 individuals vs. AS patients with predominantly controlled disease, and the larger sample size in
298 our RA group [**Methods**]), microbiome alterations across different inflammatory arthritis
299 phenotypes were, when detectable, largely shared among such subpopulations. We found many
300 of the same microbes associated with either AS (**Fig. 2**) or BASDAI (**Supplemental Fig. S6**) as
301 we identified within our RA individuals. We hypothesize this is most likely due to the consistent
302 collection, sequencing, and analysis methods applied throughout our cohort, as well as its relative
303 geographical and environmental homogeneity, any of which can otherwise cause inter-study
304 differences(*80, 81*).

305 The observation of consistent shifts in the functional capacity of gut microbial communities in
306 patients with inflammatory arthritis provides the opportunity to explain their potential chemical and
307 regulatory consequences. These include changes in folic acid metabolism, iron sequestration,
308 metabolism of broad classes of B vitamins, and production of isoprenoids (**Supplemental Fig.
309 S15-26**). Folic acid metabolism in particular (i.e. microbial processing of folate to downstream
310 compounds) was more abundant in arthritis patients with higher circulating CRP (**Fig. 4**). MTX is
311 a dihydrofolate antagonist, which competitively binds to and blocks several folate pathway
312 enzymes. In treatment-naïve RA patients, higher basal folate metabolism has been documented
313 in the peripheral serum, and MTX treatment was shown to normalize that level of folate
314 metabolism(*82*). Treatment with MTX often relieves arthritis patients of many of their joint
315 inflammation symptoms, indicating a potential role of folic acid metabolism in the disease etiology
316 (although it is not clear that this is the mechanism of action for MTX in this case). Almost no
317 patients in this study were currently taking MTX (e.g. none of the RA patients), and increased
318 microbial folic acid metabolism thus appears independent of MTX exposure. Further, several
319 studies have implicated the role of well-regulated folate metabolism in the appropriate functioning
320 of the host's immune system(*83-86*), including NK cells(*83*), the proliferation of CD8+ T
321 lymphocytes(*85*), the survival of FOX3p+ regulatory T cells(*86*). One previous study, in a much
322 smaller cohort, has also identified changes in folic acid metabolism pathways associated with
323 disease improvement within the gut ecosystem of patients with RA(*41*). Thus, folic acid
324 metabolism within the gut microbiome is a potential player in the aetiology of arthritis, and
325 warrants further mechanistic validation both linked to and independently of MTX usage.
326
327 Despite greatly increasing both the sample size and depth of microbial data compared to previous
328 studies, the inter-individual diversity of the human gut microbiome means that our results are still
329 derived from a relatively small sample size - notably from a single country and dominated by a
330 single ethnicity. This is especially true with the confounding nature of clinical data, including site
331 specific collection of distinct diagnoses, a huge age range, inherent differences in the sex
332 distribution and uneven loading of arthritis subtypes across sequencing batches. Further, since
333 we only used sequencing data, especially since these data are based on DNA profiles only, we
334 do not have a true functional profile. Thus, as noted several times above, it is impossible to
335 establish the causality or mechanism of these gut microbial changes from an observational human
336 study, and we fully expect our own and others' longitudinal human and model system research to
337 clarify these.

338 However, this study is the first to comprehensively evaluate functional changes within the gut
339 microbiome of patients with RA and AS at scale. Strikingly, we found what are becoming canonical
340 shifts in the distribution of several microbial processes in the gut during inflammation, not just in
341 local gastrointestinal conditions but during systemic inflammatory disease. Our study contributes
342 to the growing body of evidence that the gut microbiome and inflammation throughout the body
343 are tightly coupled, likely both casually and responsively, as the gut microbiome serves as a
344 mediator of environmental triggers and then also changes in response to immune activity. We
345 hypothesize that this occurs in part due to a functional "echo" of systemic inflammation in the gut

346 microbiome, due to the striking similarity in the specific processes that are altered in IBD and in
347 arthritis (to a lesser magnitude). Some of these alterations, such as those for B vitamin
348 metabolism (including both B9 and B12) could represent mechanisms for long-term prevention,
349 risk reduction, or treatment as well, as could microbial iron sequestration during arthritis-linked
350 anemia. We thus expect these results and resources to date to represent a first step in
351 understanding and managing inflammatory arthritis through its interplay with the gut microbiome.

Methods

Cohort overview

354 Subjects were recruited for this multi-center study in Birmingham, UK (primarily RA and HC
355 patients, exact numbers in **Supplemental Table 1**), Newcastle, UK (RA, PsA, NIJP, AS, and HC),
356 and Oxford, UK (AS and HC only) from June 2015 until March 2020. Patients enrolled for this
357 study were aged 17 to 97 years. As expected based on disease epidemiology(*87, 88*), diagnoses
358 were skewed by female sex, comprising 63%, 30%, 40%, 85%, and 58% of the patients with RA,
359 AS, PsA, NIJP, and HC, respectively. The majority of participants reported non-Hispanic white
360 ethnicity (74%). Approximately 50% of our population had never smoked cigarettes, and this was
361 generally lower among cases than controls (**Supplemental Table 1**).

Use of human subjects

363 All samples were collected under the Inflammatory Arthritis Microbiome Consortium umbrella
364 biospecimen protocol from one of the three main collection sites. For this study we focused on
365 adults with clearly defined arthritis (corresponding to collection sites in Oxford, Birmingham, and
366 Newcastle, UK). Patients were consented by their treating physician to have blood and stool
367 collected. Other important clinical metadata captured by the treating physician included disease-
368 specific measures of BASDAI (for AS patients), DAS28 (for RA and PsA), and swollen/active joint
369 counts. All clinical and demographic information was curated and securely housed in REDCap(*89,
370 90*). Participants provided written informed consent. The study was approved by the Newcastle
371 and North Tyneside Regional (REC 12/NE/0251), Oxfordshire (REC 06/Q1606/139) and West
372 Midlands-Back Country (REC 12/WM/0258) Research Ethics Committees.

Collection sites and clinical information

374 Biological material was obtained from consenting patients referred from primary care with
375 suspected arthritis and seen in either the Newcastle Early Arthritis Clinic (NEAC), UK [Newcastle
376 upon Tyne NHS Foundation Trust], the Birmingham Early Arthritis Cohort from Sandwell and West
377 Birmingham NHS Trust, University Hospitals Birmingham NHS Foundation Trust, and the Nuffield
378 Orthopaedic Center (NOC) and Oxford biobank during 2017-2019. The recruitment strategy was
379 designed to minimize enrolment of individuals exposed to systemic corticosteroids and/or
380 disease-modifying anti-rheumatic drugs (DMARDs) prior to biological sampling. Clinical
381 diagnoses were ascertained by board-certified rheumatologists in accordance to standard and
382 appropriate classification criteria where available. RA was assigned only where 1987 American
383 College of Rheumatology or 2010 European League Against Rheumatism/American College of
384 Rheumatology classification criteria were fulfilled. All axial spondyloarthritis patients met the
385 Assessment of Spondyloarthritis International Society (ASAS) criteria for axial
386 spondyloarthritis(*91*). When plain radiographs were performed, the vast majority (>90%)
387 additionally met modified New York Criteria for Ankylosing Spondylitis and we have hence used
388 AS. A diagnosis of non-inflammatory joint pain (NIJP) was assigned when the consulting
389 rheumatologist considered the presentation neither attributable to an inflammatory arthritis, nor to

390 osteoarthritis. For the HC's from the Oxford biobank, samples were selected to enrich for HLA-
391 B27 positive individuals, specifically selecting for ~50% of the controls being positive for the HLA-
392 B27 allele.

Sample collection

<u>Blood collection and processing</u>

395 Blood samples were collected at routine clinical visits. Whole blood was drawn and stored at –
396 80 °C. For serum, blood was drawn into a 5-ml SST tube and allowed to clot at room temperature
397 after centrifugation for 15 min at 1000G. Serum supernatants were aliquoted and stored at –80 °C
398 and in accordance with approved protocols. Blood samples were assayed for CRP, RF, anti-CCP,
399 HLA-B27 genotype, full blood count, and liver function levels by UK certified labs. RF and anti-
400 CCP were classified as positive or negative according to local laboratory cut-offs.

<u>Stool collection</u>

402 Those enrolled were provided with a previously-validated stool collection kit designed to maximize
403 ease of participation and to impart a minimum perturbative effect on downstream extraction and
404 computational protocols(*92*). They were also furnished with a brief dietary inventory modeled after
405 prior investigations and a questionnaire surveying various microbiome-relevant exposures, such
406 as the recent use of antibiotics, each completed at the time of collection.

407 All EtOH fixed kits, for metagenomic sequencing, were returned within 1-3 days of a matched
408 blood sample via Royal Mail Response Service Delivery. Participants used a FecesCatcher (Tag
409 Hemi) to collect a stool aliquot into a Sarstedt Feces container containing 100% molecular biology
410 grade ethanol (Merck Life Sciences) to facilitate preservation at more ambient temperatures.
411 Participants returned the samples enclosed in UN3373 Category B Postal kit boxes (Air Sea
412 Containers) to the Kennedy Institute of Rheumatology via Royal Mail delivery. Samples were
413 immediately stored at -80°C upon arrival until processed for DNA/RNA extractions.

414 For fresh-frozen stool, used for bacterial isolation, participants were provided with stool collection
415 kits during clinic visits. Samples were collected at home using a FecesCatcher (Tag Hemi) and a
416 feces container (Starstedt) and placed inside a disposable styrofoam container with frozen ice
417 packs. Participants returned the kits to the clinic in person and samples were frozen at -80°C
418 immediately upon arrival. Fresh-frozen stool was pulverized into a homogenous mixture utilizing
419 a Biopulverizer (Stratech) cooled in liquid nitrogen prior to aliquoting

Phenotype and clinical covariate curation

421 Study data were collected and managed using REDCap electronic data capture tools hosted at
422 The Kennedy Institute for Rheumatology(*89, 90*). REDCap (Research Electronic Data Capture)
423 is a secure, web-based software platform designed to support data capture for research studies,
424 providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data
425 manipulation and export procedures; 3) automated export procedures for seamless data
426 downloads to common statistical packages; and 4) procedures for data integration and
427 interoperability with external sources. Data dictionary for the variables included in this study can
428 found in **Supplemental Table 2**.

Microbiome sequencing

<u>DNA extraction</u>

Prior to extraction, stool aliquots were transferred to a Lysing Matrix E tube (MP Biomedicals) containing 500ul of extraction buffer consisting of 200 mM NaCl, 20 mM EDTA, 4 M guanidine thiocyanate and 1 % (v/v) β-mercaptoethanol. After adding 210ul of 20% SDS, tubes were vortexed and homogenised in a Precellys tissue homogeniser 6500 for 30 seconds. Total nucleic acid was extracted from stool samples using the Zymo ZR-duet DNA/RNA mini-kit (Zymo Research) following the manufacturer's standard instructions. DNA was eluted in DNAse/RNAse-free water and stored at -80°C.

<u>Shotgun metagenomic sequencing</u>

All sequencing was completed at the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (Oxford, UK). Briefly, DNA was first quantified using two mechanisms 1) Quant-iT^{TM} PicoGreen$^®$ dsDNA kits (Invitrogen, Carlsbad, CA, USA) and 2) a FLUOstar OPTIMA plate scanner (BMG Labtech, Offenburg, Germany); both protocols followed the manufacturers' instructions. For shotgun metagenomic sequencing, approximately 500ng of gDNA extracted from stool samples was fragmented using an Episonic system (Epigentek, Farmingdale, NY, USA). Library preparation proceeded using NEBNext Ultra DNA Sample Prep Master Mix Kit (NEB, Ipswitch, MA, USA), with some minor modifications to the base protocol and a customized automated preparation system on the Biomek FX (Beckman, Jersey City, NJ, USA). Adaptors were ligated using the Illumina (Illumina, San Diego, CA, USA) Multiplex Adaptors, and size selection proceeded with Ampure beads (Agencourt, Beckman, Jersey City, NJ, USA). Libraries were enriched with PCR. Then prepared libraries were QC'ed with Ampure beads and PicoGreen$^®$ before paired end sequencing over 8 lanes of HiSeq 4000. Each lane generated 72-98GB of data resulting in an average of 59,892,940 reads (s.d. 9,217,417) per sample.

Metagenomic analysis

<u>Bioinformatic processing</u>

All sequence processing was completed using the bioBakery shotgun metagenomic workflow v0.13.2(*93*), with components as described below. When applicable, after demultiplexing, multiple barcodes per sample were first concatenated, as all samples were run on two separate sequencing runs ~96 samples at a time (referred to here as sequencing batch).

Then samples were QC'ed for read quality and removal of human reads. Briefly, KneadData v0.7 was used to align all reads to a human and common sequencing contaminant database (DB: Homo_sapiens_hg37_and_human_contamination v0.1 and SILVA_128 DB v0.1) to remove contaminant reads from each file. Trimming and low-quality read removal proceeded via Trimmomatic(*94*) first through a sliding window trim set at a window size of 4, with an average quality score of 20 for the window. Once cut, the minimum length for the remaining sequence had to be longer than 50 base pairs. A schematic of the remaining high-quality reads is in **Supplementary Fig. S1**, with remaining average read count per sample 54,818,546 (s.d. 9,034,220).

<u>Taxonomic assignment</u>

The resulting high-quality reads were then taxonomically profiled using MetaPhlAn 2 v2.2.0(*95, 96*)(https://github.com/biobakery/MetaPhlAn). Briefly, per-species marker genes are identified by clustering all open reading frames per species (as deposited at the NCBI) into pangenomes, then

472 identifying the core genes conserved across each species, and refining these to at most 200 that
473 do not occur outside of each species. For subsequent profiling, Bowtie2(*97*) was used to align
474 each sample's reads to the resulting marker genes for taxonomic assignments. Read count per
475 marker gene is normalized by marker gene length, outliers with low or high coverage were
476 trimmed, and relative abundance was calculated per species and up the stratified taxonomy. For
477 analysis, the resulting feature tables were stratified by taxonomic level (species used unless
478 otherwise stated) and filtered for features with a minimum of 10% prevalence at 0.001%
479 abundance.

480 Functional profiling

481 Each metagenomic sample was also functionally profiled using HUMAnN 2 v0.11.1(*98*)
482 (https://github.com/biobakery/humann). Briefly, HUMAnN leverages the same per-species
483 pangenomes as MetaPhlAn to create species-specific functional databases, then aligns the reads
484 to those pangenomes again using Bowtie2(*97*). Any reads remaining after this search are then
485 translated to the corresponding amino acids and searched against UniProt(*99*) using
486 DIAMOND(*100*). This allows for the characterization of the encoded functions and the per-species
487 contribution of those functions. Once generated, HUMAnN genes were regrouped into Enzyme
488 Commission (EC) numbers and MetaCyc(*101*) metabolic pathways for analysis and converted to
489 relative abundances. For downstream analysis, features were again retained at a 10% prevalence
490 and 0.001% abundance threshold.

491 Strain profiling

492 Finally, we completed strain-level profiling through two different mechanisms: (1) SNV profiling to
493 call the dominant strain per species within each sample, using default StrainPhlAn v2.6.0(*74*)
494 parameters (http://segatalab.cibio.unitn.it/tools/strainphlan/) and (2) genotyping each sample
495 based on the presence and absence of genes within the entire species' pangenome, leveraging
496 the default settings of PanPhlAn v1.2.2.5(*75*) (http://segatalab.cibio.unitn.it/tools/panphlan/).
497 StrainPhlAn provides a single genotype for the dominant strain per species per metagenome by
498 variant-calling within the unique per-species marker genes (for species with sufficient read depth
499 to do so). PanPhlAn likewise provides a single genotype per strain per species per sample, but
500 does so by identifying which genes from the species pan-genome are present or absent in
501 sufficiently-covered organisms.

502 For both StrainPhlAn and PanPhlAn, we analyzed the top 20 most abundant species within our
503 dataset and a further curated list of 16 additional targeted (*a posteriori*) species from our MaAsLin
504 2 linear modelling results based on previous literature of interest(*27-33*). From the resulting
505 StrainPhlAn sequence genotypes, we calculated the phylogenetic distance between every pair of
506 applicable samples using ape::dist.dna function v5.3(*102*) in R, then tested for significantly
507 nonrandom phylogenetic partitioning of participant phenotypes using the adonis function within
508 vegan v2.5-6 package(*103*). We assessed signatures of positive or negative selection per strain-
509 specific marker gene genotype using the *ape::dnds* function.

510 From PanPhlAn genotypes, we removed genes either too core/prevalent or too rare to analyze,
511 i.e. greater than 80% prevalent overall or less than 30% prevalent in either the control or the
512 disease associated groupings. We further limited the UniRef families included in our analysis by
513 using a *t.test* in R for differences between the control (HCs and NIJP) and the inflammatory
514 arthritis samples (AS, RA, and PsA). Any gene with FDR q<0.2 was included in the subsequent
515 analysis. Then we calculated the Jaccard distances in a pairwise manner using the vegan v2.5-6
516 package(*103*) in R.

Bacterial isolates analysis

To isolate R. gnavus from donors stools we utilize a robotized platform that enables isolation and culturing of a high proportion of the bacteria found in the human gut(*104*). Isolation was carried out on fresh-frozen samples collected separately from ethanol-fixed samples used for sequencing (see **Methods**). Briefly, clarified stool samples are plated on solid media, followed by growth under a range of environmental conditions designed to cultivate anaerobic, microaerophilic, aerobic and spore-forming bacteria. Next, 384 colonies are picked for each donor sample and regrown in liquid media in multiwell plates. Each isolate is then identified by a combination of MALDI-TOF mass spectrometry and whole genome sequencing. Using this knowledge, the original 384 isolates are de-replicated and unique strains for each donor are archived in multiwell plates, which allows for automated selection of specific strains and sub-communities. The sequencing reads from each cultured isolated were quality filtered with Trimmomatic(*94*) and assembled using Spades(*105*). Assembly strains from patients were then included as references in StrainPhlAn.

Gnotobiotic Mouse Model of Inflammatory Arthritis

All mouse experiments were approved by the Mount Sinai Institutional Animal Care and Use Committee. Germ-free SKG mice(*106*) were bred in isolators at the Icahn School of Medicine Gnotobiotic Facility with free access to sterile water and chow (5K67, LabDiet) and under a 12-hour light/dark cycle. At 6-8 weeks of age, both male and female mice were colonized with a single strain of *R. gnavus* by oral gavage. Mice were then housed in positive pressure cages and maintained under strict aseptic conditions. Two weeks after colonization, arthritis was induced by intraperitoneal injection of 3 mg of Curdlan (beta-1,3-glucan) in PBS. Arthritis score was monitored weekly using a scoring system and by measuring the dimension of peripheral joints in three locations bilaterally: dorsal-palmer dimension of front paw, dorsal-plantar dimension of hind paw footpad, and width of hind paw 'ankle'. The change from baseline of the sum of these six measurements is reported. The scoring system used was as follows: 0; no swelling or redness, 0.5; swelling or redness of digits and mild swelling and/or redness of the joint, 1; mild swelling of the joint, 2; moderate swelling the joint, 3 substantial swelling of the joint. The sum of the score for all four paws is reported.

Downstream statistics and visualization

Statistics

Two primary classes of statistical testing were used throughout this analysis, omnibus tests and per-feature tests. The former assessed whether whole microbial community structure was significantly different based on phenotype, whereas the latter assessed this for each individual feature (e.g., taxon, pathway, etc.). Omnibus tests were generally carried out using Bray-Curtis-based PERMANOVA for the taxonomic, pathway and, enzyme-based feature tables using the vegan v2.5-6 package(*103*) in R. With one notable expectation, we ran all models in an adjusted univariate format, consisting of sequencing batch followed by the variable of interest [i.e. *adonis(bray ~ batch + x)*] with 1,000 permutations. However, to test the collection center variable, we added diagnosis to the adjustments to the model, since diagnosis was substantially confounded with clinical site (i.e., patients carrying certain diagnoses were only seen at certain centers). All p-values that are presented are FDR corrected using the *p.adjust* function unless otherwise stated.

561 Additionally, for diagnosis and inflammation status, we used the package default pairwise.adonis
562 v0.0.1(*107*) to identify which of the diagnoses or inflammation status categories might be driving
563 the overall results. The same method PERMANOVA model was used for strain testing, but with
564 Kimura 2-parameter distances as input (ape::dist.dna). For our PanPhlAn presence/absence
565 data, we tested for significantly different presence or absence of genes using a chi-Squared test
566 by leveraging the GTest function in the DescTools package(*108*) in R.

567 For parametric feature-wise multivariable testing we used MaAsLin 2 v1.4.0(*109*) in R, which finds
568 associations between microbial features and metadata of interest. MaAsLin uses a transformed
569 generalized linear model to associate each feature iteratively with covariates of interest, here
570 using a variance-stabilizing log transformation plus a small pseudocount of half the minimum
571 feature value for microbial relative abundances (total sum scaling). It then models each microbial
572 feature as a function of the patient's age and adjusts the resulting p-values for multiple hypothesis
573 tests, using BH correction. As noted above for different analyses, we used several variants of the
574 main feature-covariate model (**Equation 1**): first, a fully multivariable model that was the most
575 conservative adjusting for the most patient information; second, a reduced model accounting only
576 for technical sequencing batch; and finally, the least conservative model only adjusting for age
577 with the metadata of interest (inflammation status or patient diagnosis or anemia status). In
578 general, the model without sequencing batch was not used, since there was little evidence of
579 technical batch effects and since inflammation and diagnosis were imbalanced across our
580 sequencing batches. Further we compared the categorized way of processing CRP data to other
581 disease specific markers including BASDAI and the Das28 metrics and they identified many of
582 the same taxonomic features and importantly in the same direction (**Supplementary Fig. S6**)

583   **Equation 1:** *Feature ~ batch + drug + age + inflammation status OR patient diagnosis OR anemia status*

584 <u>Visualization</u>

585 Most visualizations were carried out using standard methods in R ggplot2 v3.3.2(*110*), ggridges
586 v0.5.2(*111*), ggthemes v4.2.0(*110*), gridExtra v2.3(*112*), gtools v3.81(*113*), and ggtree
587 v2.0.2(*114, 115*). The principal coordinate analysis was done using the capscale function in
588 vegan(*103*).

589 Additionally, we used GraPhlAn v0.9.7(*116*) to construct the cladogram in **Fig. 2**. Additional R
590 packages used for data manipulation and processing include pylr v1.8.6(*117*), dpylr v1.0.2(*118*),
591 scales v1.1.0(*118*), mgsub v1.71(*119*), and RColorBrewer v1.1.2(*120*).

Supplemental Material

593 Supplemental Results Text
594 Supplemental Discussion
595 Supplemental References
596 Supplemental Table S1: Demographic information stratified by patient diagnosis.
597 Supplemental Table S2: Data dictionary of the data reported within the manuscript.
598 Supplemental Figure S1: Quality control metrics for each sample included, or excluded from this population.
599 Supplemental Fig. S2: Additional patient demographic information.
600 Supplemental Figure S3: Community-wide alterations in the gut microbiome by patient diagnosis and inflammation
601 status.
602 Supplemental Figure S4: Consistent taxonomic alterations across the inflammatory arthritis phenotypes.
603 Supplemental Figure S5: Species-level associations with Anemia, AS, RA, inflammation and some inflammation.
604 Supplemental Figure S6: Large overlap in the taxonomic features associated with the categorized inflammation and all
605 other inflammatory markers.
606 Supplemental Figure S7: Higher order phylogenetic associations with Anemia, AS, RA, inflammation and some
607 inflammation.

608 Supplemental Figure S8: Predictive analysis largely confirms per-feature taxonomic and functional results.
609 Supplemental Figure S9: Normal abundance and prevalence of Prevotella copri in this population of patients with
610 inflammatory arthritis.
611 Supplemental Figure S10: Several species of microbes were tested for subclade level structuring within the gut
612 microbiome, only *Ruminococcus* was significantly associated with clinical features.
613 Supplemental Figure S11: Limited sub-species level associations with diagnosis.
614 Supplemental Figure S12: Further details of Ruminococcus gnavus's subclade structuring with patient diagnosis.
615 Supplemental Figure S13: Pangenome changes associated with the disease associated *Clade one* of Ruminococcus
616 gnavus.
617 Supplemental Figure S14: Entire dN/dS results from the R. gnavus marker genes.
618 Supplemental Figure S15: Differential encoding of vitamin B processing within the gut ecosystem of patients with
619 inflammatory arthritis.
620 Supplemental Figure S16: Extension of the Pathways and enzymes dysregulated in vitamin B9, folate, encoding.
621 Supplemental Figure S17: Anemia is associated with loss of encoding of vitamin B12 salvage and biosynthesis.
622 Supplemental Figure S18: Example features and the species encoding the enzymes and pathways involved in the
623 differential encoding of vitamin B processing in inflammatory arthritis.
624 Supplemental Figure S19: Downstream pathways dependent on folate metabolism and vitamin B cofactors including
625 the biosynthesis of nucleic acids and amino acids were dysregulated.
626 Supplemental Figure S20: Broadly functional features associated with cofactor, carrier, and vitamin processing were
627 deregulated in the gut ecosystem of patients with inflammatory arthritis.
628 Supplemental Figure S21: Bacteria involved in the encoding of iron sequestration were dependent on the mechanism
629 of sequestration.
630 Supplemental Figure S22: The encoding of isoprenoid biosynthesis was highly enriched in the gut ecosystem of patients
631 with inflammatory arthritis.
632 Supplemental Figure S23: Differential encoding of pathways and enzymes was observed across several broad
633 environmental response categories.
634 Supplemental Figure S24: Pathways and enzymes involved in long and short chain fatty acids biosynthesis,
635 degradation, and processing, were dysregulated.
636 Supplemental Figure S25 Aromatic, amine, and polyamine processing was differentially encoded across the
637 inflammatory and diagnosis categories.
638 Supplemental Figure S26: Broadly pathways and enzymes in sugar metabolism and glycan metabolism were
639 dysregulated in the gut ecosystems of patients with inflammatory arthritis.

References

641 1.   J. Lloyd-Price *et al.*, Multi-omics of the gut microbial ecosystem in inflammatory bowel
642   diseases. *Nature* **569**, 655-662 (2019).
643 2.   G. Nakatsu *et al.*, Gut mucosal microbiome across stages of colorectal carcinogenesis.
644   *Nat Commun* **6**, 8727 (2015).
645 3.   A. D. Kostic *et al.*, The dynamics of the human infant gut microbiome in development and
646   in progression toward type 1 diabetes. *Cell Host Microbe* **17**, 260-273 (2015).
647 4.   J. Henao-Mejia *et al.*, Inflammasome-mediated dysbiosis regulates progression of NAFLD
648   and obesity. *Nature* **482**, 179-185 (2012).
649 5.   S. Safiri *et al.*, Global, regional and national burden of rheumatoid arthritis 1990-2017: a
650   systematic analysis of the Global Burden of Disease study 2017. *Ann Rheum Dis* **78**,
651   1463-1471 (2019).
652 6.   J. T. Liu, H. M. Yeh, S. Y. Liu, K. T. Chen, Psoriatic arthritis: Epidemiology, diagnosis, and
653   treatment. *World J Orthop* **5**, 537-543 (2014).
654 7.   L. E. Dean *et al.*, Global prevalence of ankylosing spondylitis. *Rheumatology (Oxford)* **53**,
655   650-657 (2014).
656 8.   H. J. Wu *et al.*, Gut-residing segmented filamentous bacteria drive autoimmune arthritis
657   via T helper 17 cells. *Immunity* **32**, 815-827 (2010).
658 9.   G. T. Nepom *et al.*, HLA genes associated with rheumatoid arthritis. Identification of
659   susceptibility alleles using specific oligonucleotide probes. *Arthritis Rheum* **32**, 15-21
660   (1989).

661   10.   B. Mesko *et al.*, Peripheral blood gene expression patterns discriminate among chronic
662   inflammatory diseases and healthy controls and identify novel targets. *BMC Med*
663   *Genomics*, 3-15 (2010).

664   11.   A. H. M. Helm-van Mil, J. Z. Wesoly, T. W. Huizinga, Understanding the genetic
665   contribution to rheumatoid arthritis. *Curr Opin Rheumatol* **17**, 299-304 (2005).

666   12.   J. Kurko *et al.*, Genetics of rheumatoid arthritis - a comprehensive review. *Clin Rev Allergy*
667   *Immunol* **45**, 170-179 (2013).

668   13.   A. J. MacGregor *et al.*, Characterizing the quantitative genetic contribution to rheumatoid
669   arthritis using data from twins. *Arthritis Rheum* **43**, 30-37 (2000).

670   14.   D. A. Brewerton *et al.*, Ankylosing spondylitis and HL-A 27. *Lancet* **1**, 904-907 (1973).

671   15.   U. Kiltz *et al.*, Do patients with non-radiographic axial spondylarthritis differ from patients
672   with ankylosing spondylitis? *Arthritis Care Res (Hoboken)* **64**, 1415-1422 (2012).

673   16.   M. Rudwaleit *et al.*, The early disease stage in axial spondylarthritis: results from the
674   German Spondyloarthritis Inception Cohort. *Arthritis Rheum* **60**, 717-727 (2009).

675   17.   L. Klareskog, L. Padyukov, L. Alfredsson, Smoking as a trigger for inflammatory rheumatic
676   diseases. *Curr Opin Rheumatol* **19**, 49-54 (2007).

677   18.   D. D. Chaplin, Overview of the immune response. *J Allergy Clin Immunol* **125**, S3-23
678   (2010).

679   19.   G. A. Bannatyne, A. S. Wohlmann, RHEUMATOID ARTHRITIS : ITS CLINICAL
680   HISTORY, ETIOLOGY, AND PATHOLOGY. *Lancet* **147**, 1120-1125 (1896).

681   20.   J. D. Taurog *et al.*, The germfree state prevents development of gut and joint inflammatory
682   disease in HLA-B27 transgenic rats. *J Exp Med* **180**, 2359-2364 (1994).

683   21.   H. C. Rath *et al.*, Normal luminal bacteria, especially Bacteroides species, mediate chronic
684   colitis, gastritis, and arthritis in HLA-B27/human beta2 microglobulin transgenic rats. *J Clin*
685   *Invest* **98**, 945-953 (1996).

686   22.   Z. Sinkorova, J. Capkova, J. Niederlova, R. Stepankova, J. Sinkora, Commensal intestinal
687   bacterial strains trigger ankylosing enthesopathy of the ankle in inbred B10.BR (H-2(k))
688   male mice. *Hum Immunol* **69**, 845-850 (2008).

689   23.   S. Abdollahi-Roodsaz *et al.*, Stimulation of TLR2 and TLR4 differentially skews the
690   balance of T cells in a mouse model of arthritis. *J Clin Invest* **118**, 205-216 (2008).

691   24.   M. F. Van den Broek, L. B. Van de Putte, W. B. Van den Berg, Crohn's disease associated
692   with arthritis: a possible role for cross-reactivity between gut bacteria and cartilage in the
693   pathogenesis of arthritis. *Arthritis Rheum* **31**, 1077-1079 (1988).

694   25.   P. E. Pease, H. Chahal, J. E. Tallack, M. R. Lane, R. N. Allan, Cross-reactivity studies on
695   bacteria believed to be associated with inflammatory bowel disease (IBD), ankylosing
696   spondylitis (AS) and reactive arthritis (ReA). *Br J Rheumatol* **27 Suppl 2**, 32-33 (1988).

697   26.   S. N. Lichtman *et al.*, Reactivation of arthritis induced by small bowel bacterial overgrowth
698   in rats: role of cytokines, bacteria, and bacterial polymers. *Infect Immun* **63**, 2295-2301
699   (1995).

700   27.   E. Eerola *et al.*, Intestinal flora in early rheumatoid arthritis. *Br J Rheumatol* **33**, 1030-1038
701   (1994).

702   28.   J. Vaahtovuo, E. Munukka, M. Korkeamaki, R. Luukkainen, P. Toivanen, Fecal microbiota
703   in early rheumatoid arthritis. *J Rheumatol* **35**, 1500-1505 (2008).

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

706   30.   S. Stebbings *et al.*, Comparison of the faecal microflora of patients with ankylosing
707   spondylitis and controls using molecular methods of analysis. *Rheumatology (Oxford)* **41**,
708   1395-1401 (2002).

709   31.   M. E. Costello *et al.*, Brief Report: Intestinal Dysbiosis in Ankylosing Spondylitis. *Arthritis*
710   *Rheumatol* **67**, 686-691 (2015).

711 32. M. L. Stoll *et al.*, Altered microbiota associated with abnormal humoral immune responses
712   to commensal organisms in enthesitis-related arthritis. *Arthritis Res Ther* **16**, 486 (2014).
713 33. J. U. Scher *et al.*, Decreased bacterial diversity characterizes the altered gut microbiota in
714   patients with psoriatic arthritis, resembling dysbiosis in inflammatory bowel disease.
715   *Arthritis Rheumatol* **67**, 128-139 (2015).
716 34. X. Zhang *et al.*, The oral and gut microbiomes are perturbed in rheumatoid arthritis and
717   partly normalized after treatment. *Nat Med* **21**, 895-905 (2015).
718 35. C. Wen *et al.*, Quantitative metagenomics reveals unique gut microbiome biomarkers in
719   ankylosing spondylitis. *Genome Biol* **18**, 142 (2017).
720 36. R. Huang *et al.*, Metagenome-wide association study of the alterations in the intestinal
721   microbiome composition of ankylosing spondylitis patients and the effect of traditional and
722   herbal treatment. *J Med Microbiol* **69**, 797-805 (2020).
723 37. J. Yin *et al.*, Shotgun metagenomics reveals an enrichment of potentially cross-reactive
724   bacterial epitopes in ankylosing spondylitis patients, as well as the effects of TNFi therapy
725   upon microbiome composition. *Ann Rheum Dis* **79**, 132-140 (2020).
726 38. C. Zhou *et al.*, Metagenomic profiling of the pro-inflammatory gut microbiota in ankylosing
727   spondylitis. *J Autoimmun* **107**, 102360 (2020).
728 39. T. Kishikawa *et al.*, Metagenome-wide association study of gut microbiome revealed novel
729   aetiology of rheumatoid arthritis in the Japanese population. *Ann Rheum Dis* **79**, 103-111
730   (2020).
731 40. M. Berland *et al.*, Both Disease Activity and HLA-B27 Status Are Associated With Gut
732   Microbiome Dysbiosis in Spondyloarthritis Patients. *Arthritis Rheumatol*, (2022).
733 41. V. K. Gupta *et al.*, Gut microbial determinants of clinically important improvement in
734   patients with rheumatoid arthritis. *Genome Med* **13**, 149 (2021).
735 42. J. U. Scher, D. R. Littman, S. B. Abramson, Microbiome in Inflammatory Arthritis and
736   Human Rheumatic Diseases. *Arthritis Rheumatol* **68**, 35-45 (2016).
737 43. Y. Zhao *et al.*, Hidden link in gut-joint axis: gut microbes promote rheumatoid arthritis at
738   early stage by enhancing ascorbate degradation. *Gut* **71**, 1041-1043 (2022).
739 44. H. W. van Steenbergen *et al.*, EULAR definition of arthralgia suspicious for progression to
740   rheumatoid arthritis. *Ann Rheum Dis* **76**, 491-496 (2017).
741 45. M. M. Newkirk, A. Zbar, M. Baron, A. R. Manges, Distinct bacterial colonization patterns
742   of Escherichia coli subtypes associate with rheumatoid factor status in early inflammatory
743   arthritis. *Rheumatology (Oxford)* **49**, 1311-1316 (2010).
744 46. A. Pianta *et al.*, Evidence of the Immune Relevance of Prevotella copri, a Gut Microbe, in
745   Patients With Rheumatoid Arthritis. *Arthritis Rheumatol* **69**, 964-975 (2017).
746 47. A. Kojima *et al.*, Infection of specific strains of Streptococcus mutans, oral bacteria,
747   confers a risk of ulcerative colitis. *Sci Rep* **2**, 332 (2012).
748 48. G. Kaci *et al.*, Anti-inflammatory properties of Streptococcus salivarius, a commensal
749   bacterium of the oral cavity and digestive tract. *Appl Environ Microbiol* **80**, 928-934 (2014).
750 49. J. U. Scher *et al.*, Periodontal disease and the oral microbiota in new-onset rheumatoid
751   arthritis. *Arthritis Rheum* **64**, 3083-3094 (2012).
752 50. L. Maier *et al.*, Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*
753   **555**, 623-628 (2018).
754 51. F. Imhann *et al.*, Proton pump inhibitors affect the gut microbiome. *Gut* **65**, 740-748 (2016).
755 52. P. de Pablo, T. Dietrich, T. E. McAlindon, Association of periodontal disease and tooth
756   loss with rheumatoid arthritis in the US population. *J Rheumatol* **35**, 70-76 (2008).
757 53. R. Martin *et al.*, Faecalibacterium prausnitzii prevents physiological damages in a chronic
758   low-grade inflammation murine model. *BMC Microbiol* **15**, 67 (2015).
759 54. C. T. Brown *et al.*, Gut microbiome metagenomics analysis suggests a functional model
760   for the development of autoimmunity for type 1 diabetes. *PLoS One* **6**, e25792 (2011).

761 55. K. Machiels *et al.*, A decrease of the butyrate-producing species Roseburia hominis and
762 Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* **63**,
763 1275-1283 (2014).
764 56. B. J. Parker, P. A. Wearsch, A. C. M. Veloo, A. Rodriguez-Palacios, The Genus Alistipes:
765 Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health.
766 *Front Immunol* **11**, 906 (2020).
767 57. P. Louis, P. Young, G. Holtrop, H. J. Flint, Diversity of human colonic butyrate-producing
768 bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environ*
769 *Microbiol* **12**, 304-314 (2010).
770 58. S. L. La Rosa *et al.*, The human gut Firmicute Roseburia intestinalis is a primary degrader
771 of dietary beta-mannans. *Nat Commun* **10**, 905 (2019).
772 59. A. Picchianti-Diamanti *et al.*, Analysis of Gut Microbiota in Rheumatoid Arthritis Patients:
773 Disease-Related Dysbiosis and Modifications Induced by Etanercept. *Int J Mol Sci* **19**,
774 (2018).
775 60. R. Bodkhe, B. Balakrishnan, V. Taneja, The role of microbiome in rheumatoid arthritis
776 treatment. *Ther Adv Musculoskelet Dis* **11**, 1759720x19844632 (2019).
777 61. J. Chen *et al.*, An expansion of rare lineage intestinal microbes characterizes rheumatoid
778 arthritis. *Genome Med* **8**, 43 (2016).
779 62. M. Zhang *et al.*, Faecalibacterium prausnitzii inhibits interleukin-17 to ameliorate colorectal
780 colitis in rats. *PLoS One* **9**, e109146 (2014).
781 63. H. M. Hamer *et al.*, Review article: the role of butyrate on colonic function. *Aliment*
782 *Pharmacol Ther.* **27**, 104-119 (2008).
783 64. S. Yuille, N. Reichardt, S. Panda, H. Dunbar, I. E. Mulder, Human gut bacteria as potent
784 class I histone deacetylase inhibitors in vitro through production of butyric acid and valeric
785 acid. *PLoS One* **13**, e0201073 (2018).
786 65. X. C. Morgan *et al.*, Dysfunction of the intestinal microbiome in inflammatory bowel
787 disease and treatment. *Genome Biol* **13**, R79 (2012).
788 66. K. Atarashi *et al.*, Ectopic colonization of oral bacteria in the intestine drives T(H)1 cell
789 induction and inflammation. *Science* **358**, 359-365 (2017).
790 67. D. Azzouz *et al.*, Lupus nephritis is linked to disease-activity associated expansions and
791 immunity to a gut commensal. *Ann Rheum Dis* **78**, 947-956 (2019).
792 68. H. H. Chua *et al.*, Intestinal Dysbiosis Featuring Abundance of Ruminococcus gnavus
793 Associates With Allergic Diseases in Infants. *Gastroenterology* **154**, 154-167 (2018).
794 69. E. Coletto *et al.*, The role of the mucin-glycan foraging Ruminococcus gnavus in the
795 communication between the gut and the brain. *Gut Microbes* **14**, 2073784 (2022).
796 70. M. Breban *et al.*, Faecal microbiota study reveals specific dysbiosis in spondyloarthritis.
797 *Ann Rheum Dis* **76**, 1614-1622 (2017).
798 71. A. B. Hall *et al.*, A novel Ruminococcus gnavus clade enriched in inflammatory bowel
799 disease patients. *Genome Med* **9**, 103 (2017).
800 72. M. T. Henke *et al.*, Ruminococcus gnavus, a member of the human gut microbiome
801 associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc Natl*
802 *Acad Sci U S A* **116**, 12672-12677 (2019).
803 73. M. Schirmer, A. Garner, H. Vlamakis, R. J. Xavier, Microbial genes and pathways in
804 inflammatory bowel disease. *Nat Rev Microbiol* **17**, 497-511 (2019).
805 74. D. T. Truong, A. Tett, E. Pasolli, C. Huttenhower, N. Segata, Microbial strain-level
806 population structure and genetic diversity from metagenomes. *Genome Res* **27**, 626-638
807 (2017).
808 75. M. Scholz *et al.*, Strain-level microbial epidemiology and population genomics from
809 shotgun metagenomics. *Nat Methods* **13**, 435-438 (2016).
810 76. M. T. Henke *et al.*, Capsular polysaccharide correlates with immune response to the
811 human gut microbe Ruminococcus gnavus. *Proc Natl Acad Sci U S A* **118**, (2021).

812 77. D. O. Kennedy, B Vitamins and the Brain: Mechanisms, Dose and Efficacy--A Review.
813   *Nutrients* **8**, 68 (2016).
814 78. Q. Wang, R. Xu, Data-driven multiple-level analysis of gut-microbiome-immune-joint
815   interactions in rheumatoid arthritis. *BMC Genomics* **20**, 124 (2019).
816 79. D. Yu *et al.*, The Gut Microbiome and Metabolites Are Altered and Interrelated in Patients
817   With Rheumatoid Arthritis. *Front Cell Infect Microbiol* **11**, 763507 (2021).
818 80. R. Sinha *et al.*, Assessment of variation in microbial community amplicon sequencing by
819   the Microbiome Quality Control (MBQC) project consortium. *Nat Biotechnol* **35**, 1077-1086
820   (2017).
821 81. S. Ma *et al.*, Population structure discovery in meta-analyzed microbial communities and
822   inflammatory bowel disease using MMUPHin. *Genome Biol* **23**, 208 (2022).
823 82. M. Blits *et al.*, Methotrexate normalizes up-regulated folate pathway genes in rheumatoid
824   arthritis. *Arthritis Rheum* **65**, 2791-2802 (2013).
825 83. A. L. Bayer, C. A. Fraker, The Folate Cycle As a Cause of Natural Killer Cell Dysfunction
826   and Viral Etiology in Type 1 Diabetes. *Front Endocrinol (Lausanne)* **8**, 315 (2017).
827 84. D. N. Meadows, R. H. Bahous, A. F. Best, R. Rozen, High Dietary Folate in Mice Alters
828   Immune Response and Reduces Survival after Malarial Infection. *PLoS One* **10**,
829   e0143738 (2015).
830 85. C. Courtemanche, I. Elson-Schwab, S. T. Mashiyama, N. Kerry, B. N. Ames, Folate
831   deficiency inhibits the proliferation of primary human CD8+ T lymphocytes in vitro. *J
832   Immunol* **173**, 3186-3192 (2004).
833 86. M. Kinoshita *et al.*, Dietary folic acid promotes survival of Foxp3+ regulatory T cells in the
834   colon. *J Immunol* **189**, 2869-2878 (2012).
835 87. T. K. Kvien, T. Uhlig, S. Odegard, M. S. Heiberg, Epidemiological aspects of rheumatoid
836   arthritis: the sex ratio. *Ann N Y Acad Sci* **1069**, 212-222 (2006).
837 88. A. T. Masi, W. R. Wilkins, Does male:female sex ratio in ankylosing spondylitis change
838   with age? *J Rheumatol* **23**, 947-948 (1996).
839 89. P. A. Harris *et al.*, Research electronic data capture (REDCap)--a metadata-driven
840   methodology and workflow process for providing translational research informatics
841   support. *J Biomed Inform* **42**, 377-381 (2009).
842 90. P. A. Harris *et al.*, The REDCap consortium: Building an international community of
843   software platform partners. *J Biomed Inform* **95**, 103208 (2019).
844 91. M. Rudwaleit *et al.*, The Assessment of SpondyloArthritis International Society
845   classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general.
846   *Ann Rheum Dis* **70**, 25-31 (2011).
847 92. E. A. Franzosa *et al.*, Relating the metatranscriptome and metagenome of the human gut.
848   *Proc Natl Acad Sci U S A* **111**, E2329-2338 (2014).
849 93. L. J. McIver *et al.*, bioBakery: a meta'omic analysis environment. *Bioinformatics* **34**, 1235-
850   1237 (2018).
851 94. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence
852   data. *Bioinformatics* **30**, 2114-2120 (2014).
853 95. N. Segata *et al.*, Metagenomic microbial community profiling using unique clade-specific
854   marker genes. *Nat Methods* **9**, 811-814 (2012).
855 96. D. T. Truong *et al.*, MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat
856   Methods* **12**, 902-903 (2015).
857 97. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,
858   357-359 (2012).
859 98. E. A. Franzosa *et al.*, Species-level functional profiling of metagenomes and
860   metatranscriptomes. *Nat Methods* **15**, 962-968 (2018).
861 99. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-d515 (2019).

862   100.   B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND.
863   *Nat Methods* **12**, 59-60 (2015).
864   101.   R. Caspi *et al.*, The MetaCyc database of metabolic pathways and enzymes and the
865   BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **42**, (2014).
866   102.   E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and
867   evolutionary analyses in R. *Bioinformatics* **35**, 526-528 (2019).
868   103.   P. Dixon, VEGAN, a package of R functions for community ecology. *Journal of Vegetation
869   Science*, (2003).
870   104.   G. J. Britton *et al.*, Microbiotas from Humans with Inflammatory Bowel Disease Alter the
871   Balance of Gut Th17 and RORγt+ Regulatory T Cells and Exacerbate Colitis in Mice.
872   *Immunity* **50**, 212-224 (2019).
873   105.   A. Bankevich *et al.*, SPAdes: A New Genome Assembly Algorithm and Its Applications to
874   Single-Cell Sequencing. *J Comput Biol* **19**, 455-477 (2012).
875   106.   N. Sakaguchi *et al.*, Altered thymic T-cell selection due to a mutation of the ZAP-70 gene
876   causes autoimmune arthritis in mice. *Nature* **426**, 454-460 (2003).
877   107.   P. M. Arbizu, pairwiseAdonis: Pairwise Multilevel Comparison using Adonis. . *R package
878   version 0.0.1.*, (2017).

879   108.   A. Signorell, e. m. al., DescTools: Tools for descriptive statistics. *R package version
880   0.99.34.*, (2020).
881   109.   H. Mallick *et al.*, Multivariable Association Discovery in Population-scale Meta-omics
882   Studies. *PLoS Comput Biol* **17**, e1009442 (2021).
883   110.   J. B. Arnold, ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. *R package version
884   4.2.0*, (2019).
885   111.   C. O. Wilke, ggridges: Ridgeline Plots in 'ggplot2'. *R package version 0.5.2*, (2020).
886   112.   B. Auguie, gridExtra: Miscellaneous Functions for "Grid" Graphics. *R package version
887   2.3.*, (2017).
888   113.   G. R. Warnes, B. Bolker, T. Lurnley, gtools: Various R Programming Tools. *R package
889   version 3.8.1.*, (2018).
890   114.   G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Lam, ggtree: an r package for visualization and
891   annotation of phylogenetic trees with their covariates and other associated data. *Methods
892   in Ecology and Evolution* **8**, 28-36 (2016).
893   115.   G. Yu, T. T. Lam, H. Zhu, Y. Guan, Two Methods for Mapping and Visualizing Associated
894   Data on Phylogeny Using Ggtree. *Mol Biol Evol* **35**, 3041-3043 (2018).
895   116.   F. Asnicar, G. Weingart, T. L. Tickle, C. Huttenhower, N. Segata, Compact graphical
896   representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**, (2015).
897   117.   H. Wickham, The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical
898   Software* **40**, (2011).
899   118.   H. Wickham, R. Francois, L. Henry, K. Muller, dplyr: A Grammar of Data Manipulation. .
900   *R package version 1.0.2*, (2020).

901   119.   M. Ewing, mgsub: Safe, Multiple, Simultaneous String Substitution. . *R package version
902   1.7.1*, (2019).

903   120.   Neuwirth, RColorBrewer: ColorBrewer Palettes. *R package version 1.1-2.*, (2014).
904   121.   J. W. Kim *et al.*, Influence of proton pump inhibitor or rebamipide use on gut microbiota of
905   rheumatoid arthritis patients. *Rheumatology (Oxford)* **60**, 708-716 (2021).
906   122.   Y. Sun *et al.*, Characteristics of Gut Microbiota in Patients With Rheumatoid Arthritis in
907   Shanghai, China. *Front Cell Infect Microbiol* **9**, 369 (2019).
908   123.   T. Ganz, E. Nemeth, Iron sequestration and anemia of inflammation. *Semin Hematol* **46**,
909   387-393 (2009).

910 124. J. Kruger *et al.*, Effect of phytate reduction of sorghum, through genetic modification, on
911 iron and zinc availability as assessed by an in vitro dialysability bioaccessibility assay,
912 Caco-2 cell uptake assay, and suckling rat pup absorption model. *Food Chem* **141**, 1019-
913 1025 (2013).

914

915 **Figure captions**

916 **Figure 1: The gut microbiome is significantly altered in patients with inflammatory arthritis. A)** Overview of the
917 participants and samples collected from each subtype of arthritis and the **B)** collection schema. **C)** Prevalence of
918 arthritis subtypes by age collected under this cross-sectional study by the Inflammatory Arthritis Microbiome Consortium
919 (IAMC). **D)** Proportion of patients with overt inflammation, some inflammation and no inflammation defined by tertiles
920 of circulating serum levels of CRP by diagnosis (0-4mg/L *Not inflamed, 4-10mg/L Some inflammation, 10-167mg/L*
921 *Inflammation*). **E)** Univariate PERMANOVA of Bray-Curtis dissimilarity by demographic and clinical measures. Color
922 indicates the features assessed (microbial taxonomy, metagenomic pathways, and metagenomic Enzyme Commission
923 (ECs) numbers). Stars indicate FDR-corrected *q*-values. Tests are batch adjusted (i.e. Bray-Curtis ~ sequencing batch
924 + covariate). **The effect of the collection site was also adjusted for by diagnosis to account for the fact that some sites
925 only enrolled individuals with a particular disease subtype. **F)** Bray-Curtis principal coordinates analysis of all 440
926 taxonomic profiles, with overall diagnosis tracking weakly with a typical Bacteroidetes/Firmicutes population gradient.
927 Species are labeled by using weighted averages and limited to just species whose abundance explains samples that
928 fall outside of 0.04 distance from origin.

929 **Figure 2: Taxonomic features differ in both a diagnosis- and severity-specific manner.** Many features of the gut
930 microbiome were found to be significantly associated with the diagnosis of the patients, their current levels of systemic
931 inflammation (categorized CRP value), or their concurrent hemoglobin levels (linear model; FDR q-value < 0.25). **A)**
932 Clade-specific taxonomic alterations of the gut microbiome by inflammation (inflamed vs. not inflamed, some
933 inflammation vs. not inflamed), diagnosis (RA vs. HC, AS vs. HC), and hemoglobin levels (Anemia vs. Normal). These
934 variables have been reduced for clarity, all results are presented in **Supplemental Fig. S4**. **B)** Microbial species
935 associated with higher levels of systemic inflammation as measured by CRP in the serum. Changes were observed in
936 both the profiles of each clade's relative abundance within the gut community and, in many cases, their prevalence as
937 well. Increased abundance and prevalence were observed in three previously inflammation-associated species(*45,*
938 *121, 122*): *Streptococcus vestibularis* (linear model; coef = 0.707, FDR q-value = 0.002)(a typically oral clade), and the
939 proinflammatory *Escherichia coli* (linear model; coef = 0.788, FDR q-value = 0.006) and *Ruminococcus gnavus* (linear
940 model; coef = 1.005, FDR q-value = 0.003). **C)** Example disease-specific alterations in the gut microbiomes of patients
941 with inflammatory arthritis. *Faecalibacterium prausnitzii* exhibits a lower abundance and prevalence in patients with RA
942 (linear model; coef = -0.299, FDR q-value = 0.126) and, to a lesser extent, AS (linear model; coef = -0.344, FDR q-
943 value = 0.21), which parallels its changes in other inflammatory conditions. The typically oral-sourced clade
944 *Streptococcus salivarius* had a higher prevalence and abundance especially in RA patients. **D)** A loss of several typical
945 human gut residents was observed with lowering hemoglobin concentrations (d/L), with the highest effect size observed
946 in *Eubacterium rectale (*linear model; coef = 0.832, FDR q-value = 0.006*)*.

947 **Figure 3: Identification of a *Ruminococcus gnavus* sub-species clade enriched in patients with inflammatory**
948 **arthritis. A)** Species tested for sub-species phylogenetic structure associated with diagnosis and inflammation. Of the
949 25 taxa chosen for testing based on abundance and significant species-level disease associations, only 21 species
950 had sufficient sample sizes to investigate for disease and inflammatory associations. Only *Ruminococcus gnavus*
951 exhibited significant subspecies level structuring (Kimura 2-parameter distance, PERMANOVA; R2 = 0.18 , q-val =
952 0.01; StrainPhlAn(*65*)) (**Supplemental Fig. S10-12**) **B)** Phylogeny of *R. gnavus* strains from each individual's gut
953 microbiomes. Clade One was significantly associated (as above) with patient diagnosis, predominantly carried by AS
954 patients and a subset of RA patients. *Clade One* also contained isolates from IBD patients that were sourced from a
955 subset of the PRISM cohort(*66*) and isolates from an RA patient (**Methods**). Triangles indicate isolated strains used in
956 (D) and (E). **C)** Clade One *R. gnavus* induces an arthritis-like phenotype in SKG mice. Previously germ-free mice were
957 inoculated with a monoculture of a single *R. gnavus* strain (arrows on **Fig. 3B**) at week 1. Two to three weeks later
958 curdlan was injected to simulate arthritis. Male and female mice were then followed for joint diameter changes (**D**) and
959 disease score (**E**). Mice inoculated with the human disease-associated clade were more likely to develop severe
960 symptoms than mice exposed to other *R. gnavus* clades.

961 **Figure 4: Differential encoding of vitamin B metabolism and processing in the gut microbiomes of patients**
962 **with inflammatory arthritis. A)** Significant associations between patient diagnosis, inflammation status, and
963 hemoglobin levels with folic acid metabolism, vitamin B12 salvage and biosynthesis and methionine biosynthesis and
964 cycling (FDR q-value < 0.25). Folic acid metabolism (top; B9) shows the dichotomy of the association between upstream

965 pathway components positively associated with both diagnosis and inflammation, but the methionine component of the
966 pathway is mainly negatively associated with diagnosis (RA and AS), hemoglobin, and especially inflammation. All
967 vitamin B12 pathways and enzymes identified (bottom) were negatively associated with diagnosis, inflammation, and
968 hemoglobin levels. **B)** Folic acid metabolism pathway components carried by gut microbes. Enzymes in the one-carbon
969 metabolism cycle were enriched in gut metagenomes during inflammatory arthritis (red), while enzymes in the
970 methionine cycle were decreased (blue). Cofactors in each cycle (various B vitamins) displayed diverse enrichment
971 patterns. **C)** Increased encoding of folate metabolism cycle components within the gut microbiomes of patients with
972 inflammatory arthritis. **D)** Decreased encoding of both the salvage and biosynthesis of vitamin B12 in arthritis gut
973 ecosystems.

974 **Figure 5: Increased metagenomic carriage of pathways and enzymes involved in iron sequestration across**
975 **several distinct mechanisms in arthritis patients with high serum C-reactive protein.** Disruptions in iron
976 sequestration has been previously observed in several immune cell types in patients with inflammatory arthritis(*123*);
977 parallel functional shifts were observed here in the gut microbiome, with increased carriage of several enzyme
978 categories involved in the sequestration of iron. **A)** Metagenomic functional features (pathways and enzymes) with
979 disrupted, mainly increased, carriage in the gut ecosystems of patients with elevated C-reactive protein values.
980 Mechanisms of increased microbial iron sequestration spanned heme and non-heme (ferritin) based storage, high
981 affinity siderophores, and the production of phytates (which decrease the absorption of iron by hosts(*124*)). **B)** Enzymes
982 from several of the sequestration mechanisms. **C)** Species that most contribute to carriage of the ferrochelatase iron
983 sequestration gene family. A clear shift in species encoding this gene between not inflamed and inflamed gut
984 ecosystems, with increases coming from clades including *Escherichia coli* and *Streptococcus salivarius.*

Acknowledgements

Author contributions

1002 This study was conceptualized and designed by FP, CH, KR, PB, AGP, DRL, SPY, JJF, LRW
1003 and PCT. Data and specimen collection was completed by PC, SJB, AF, CP, SK, AGP, PB, and
1004 KR. Molecular methods prior to sequencing were done by PC, IB, CP, CM, VK, and LHL.
1005 Computational storage of clinical and sequence data was designed, maintained, and organized
1006 by KSB, NEI, PC, LJM, MAJ, and BM. Analysis was completed by KNT, KSB, NEI, GJB, SM, and
1007 LHN. Feedback and directions were guided by AGP, PB, and KR. Mouse work was completed by
1008 GJB and JJF. Manuscript was written by KNT, LHN, AP and CH. Editing was completed by all
1009 authors.

Competing interests

1011 D.R.L. is a founder and advisor to Vedanta Biosciences.

Data availability

Sequence data and metadata are available for approved user to download through the EGA (https://ega-archive.org/), study accession number EGAS00001005525. While processed taxonomic and functional tables are available in supplement (**Supplemental tables S3-5**).

Code availability

Bioinformatic workflows for metagenomic processing are available at https://huttenhower.sph.harvard.edu/biobakery_workflows, these include some basic statistical and visualization scripts. Custom analysis scripts are available at http://huttenhower.sph.harvard.edu/Adult_cross-sectional_IAMC