

Basic science

# Treatment response in rheumatoid arthritis is predicted by the microbiome: a large observational study in UK DMARD-naive patients

Nathan P. Danckert <sup>1,\*</sup>, Maxim B. Freidin <sup>2,‡</sup>, Isabelle Granville Smith <sup>1</sup>,  
Philippa M. Wells <sup>3,4</sup>, Maryam Kazemi Naeini <sup>1</sup>, Alessia Visconti <sup>1</sup>, Roger Compte <sup>1</sup>,  
Alexander MacGregor <sup>5,6</sup>, Frances M. K. Williams <sup>1,7</sup>

<sup>1</sup>Department of Twin Research and Genetic Epidemiology, School of Life Course Population Sciences, King's College London, London, UK

<sup>2</sup>Department of Biology, School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK

<sup>3</sup>UK Dementia Research Institute, Imperial College London, London, UK

<sup>4</sup>Department of Brain Sciences, Imperial College London, London, UK

<sup>5</sup>Norwich Medical School, University of East Anglia, Norwich, UK

<sup>6</sup>Rheumatology Department, Norfolk and Norwich University Hospitals NHS Trust, Norwich, UK

<sup>7</sup>Guy's and St Thomas' NHS Trust, London, UK

\*Correspondence to: Nathan Danckert, Department of Twin Research and Genetic Epidemiology, 3rd and 4th Floor, Block D, South Wing, St Thomas' Hospital, Westminster Bridge Rd, London SE1 7EH, UK. E-mail: nathan.danckert@kcl.ac.uk

‡N.P.D. and M.B.F. contributed equally.

## Abstract

**Objectives:** Disease-modifying antirheumatic drugs (DMARDs) are a first-line treatment in rheumatoid arthritis (RA). Treatment response to DMARDs is patient-specific, dose efficacy is difficult to predict and long-term results are variable. The gut microbiota are known to play a pivotal role in prodromal and early-disease RA, manifested by *revotella* spp. enrichment. The clinical response to therapy may be mediated by microbiota, and large-scale studies assessing the microbiome are few. This study assessed whether microbiome signals were associated with, and predictive of, patient response to DMARD treatment. Accurate early identification of those who will respond poorly to DMARD therapy would allow selection of alternative treatment (e.g. biologic therapy) and potentially improve patient outcome.

**Methods:** A multicentre, longitudinal, observational study of stool- and saliva microbiome was performed in DMARD-naive, newly diagnosed RA patients during introduction of DMARD treatment. Clinical data and samples were collected at baseline ( $n = 144$ ) in DMARD-naive patients and at six weeks ( $n = 117$ ) and 12 weeks ( $n = 95$ ) into DMARD therapy. Samples collected ( $n = 365$  stool,  $n = 365$  saliva) underwent shotgun sequencing. Disease activity measures were collected at each timepoint and minimal clinically important improvement determined.

**Results:** In total, 26 stool microbes were found to decrease in those manifesting a minimal clinically important improvement. *revotella* spp. and *Streptococcus* spp. were the predominant taxa to decline following six weeks and 12 weeks of DMARDs, respectively. Furthermore, baseline microbiota of DMARD-naive patients were indicative of future response.

**Conclusion:** DMARDs appear to restore a perturbed microbiome to a eubiotic state. Moreover, microbiome status can be used to predict likelihood of patient response to DMARD.

**Keywords** gut microbiome, rheumatoid arthritis, disease-modifying antirheumatic drugs, shotgun metagenomic sequencing, minimal clinically important improvement.

## Rheumatology key messages

- RA microbiota differentiates response to DMARDs.
- *revotella* spp. reduced in responders following six weeks treatment.
- *Streptococcus* spp. reduced in responders following 12 weeks.
- DMARDs appear to restore a perturbed microbiome to a more eubiotic state.

Received: 4 August 2023. Accepted: 22 December 2023

© The Author(s) 2024. Published by Oxford University Press on behalf of the British Society for Rheumatology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

Rheumatoid arthritis (RA) is an autoimmune, multi-system inflammatory disorder usually treated with disease-modifying antirheumatic drugs (DMARDs) and biologic therapy. Known to be substantially heritable [1], aetiology is a complex interaction of genetic and environmental triggers, not wholly defined. The autoimmune pathogenesis of RA takes place over many years prior to symptom onset [2]. The importance of the microbiota in the development and maintenance immune system function is well recognised [3]. Aberrant immune responsiveness may result from an abnormal microbiome and there is intense research activity to determine whether gut and oral microbes play a role in RA pathogenesis [4].

We previously used an RA polygenic risk score (PRS) in non-affected individuals to show host genetic makeup associates with enrichment for *Prevotella* spp. in the stool of TwinsUK participants who had higher PRS for RA [5]. PRS minimizes confounding by RA and treatment in association studies. Other groups have shown RA patients manifest high levels of *Prevotella copri* (*P. copri*) in stool early in symptom onset [6, 7]. Furthermore, serum anti-*P. copri* antibodies and *P. copri*-derived proteins were detectable in joint synovial fluid of RA patients [8].

DMARDs are currently the first choice of treatment for patients diagnosed with RA in the UK [9]. In some patients, disease activity is reduced, joint damage viewed radiographically is slowed, and remission may be achieved [10]. Methotrexate is the most prescribed DMARD; it inhibits multiple inflammatory pathways that are active in RA [9]. Orally administered methotrexate is metabolised by the gut microbiome and individuals' variability of microbiome abundance and diversity likely contributes to differences in medication efficacy [9]. Treatment strategies in RA currently involve introducing DMARDs at a low dose with escalation over months. Methotrexate monotherapy results in less than half of patients responding well to their prescription [11]. In patients non-responsive to first-line therapy, the iterative process of drug assaying means time is lost taking DMARD medication to which some patients never respond well, irrespective of dose. As early and aggressive therapy (i.e. treat-to-target) improves long-term outcomes, advances in patient stratification towards therapies from which they will benefit are welcome.

Predicting patient response to DMARD therapy via the microbiome would be invaluable for prognosis. Three metagenome studies have reported microbiota predicted DMARD response [12–14], although further validation is required before clinical application. In the present study, we sought to determine whether the microbiome (oral and gut) before treatment in RA predicted response to DMARDs. In this longitudinal, observational, multicentre study we recruited newly diagnosed RA patients and obtained oral and gut samples at DMARD-naïve baseline and at follow-up six weeks and 12 weeks later. We examined metagenomes to determine whether microbiome composition and functionality of DMARD-naïve patients early during RA were predictive of responses to treatment. Furthermore, we compared the microbiota composition of newly diagnosed RA patients with long-term DMARD-treated RA participants from the Norfolk Arthritis Register (NOAR).

## Materials and methods

### Study design and participants

The Influence of Methotrexate on the Rheumatoid Arthritis Microbiome (IMRABIOME) study is a longitudinal, observational study of microbiota observed in DMARD-naïve newly diagnosed RA patients who had inflammatory arthritis symptoms for <12 months and fulfilled the 2010 European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) classification criteria for RA [15]. Recruitment was undertaken between April 2017 and July 2019 from 12 outpatient rheumatology clinics in England, UK. All participants were DMARD-naïve at recruitment but due to commence DMARD treatment, prescribed by their consultant rheumatologist in accordance with UK standard treatment and National Institute for Health and Care Excellence (NICE) guidelines. Inclusion criteria comprised: age of 18 or over, with new-onset RA and having received rheumatologist advice to commence DMARD therapy. All interventions and medications were recorded in the study database including corticosteroid treatment (prednisolone <20 mg per day) and non-steroidal anti-inflammatory drugs (NSAIDs). Concurrent non-rheumatic disorders and their treatment were recorded. Exclusion criteria included previous exposure to DMARD or biologic therapy; corticosteroid treatment equivalent to prednisolone  $\geq 20$  mg/day, significant comorbidities (e.g. severe congestive heart failure; renal, hepatic or malignant disease); pregnancy; and enrolment in a clinical trial.

DMARD-naïve participants were assessed at baseline and at six and 12 weeks after commencing DMARD therapy. Intestinal (stool) and saliva samples clinical markers and relevant clinical, demographic and lifestyle information were collected. Participants completed a variety of health, well-being and activities of daily functioning questionnaires. Written informed consent was obtained from all participants and the Declaration of Helsinki (1996) was followed. Study approval was granted by the Health Research Authority (HRA) Research Ethics Committee (REC) IRAS, project 212604. Sample collection is detailed in the [Supplementary Methods](#), available at *Rheumatology* online.

### Microbiome profiling and collection

Shotgun metagenomic analysis was performed on stool and saliva samples collected from study participants, as described in the [Supplementary Methods](#), available at *Rheumatology* online. In short, samples were sent to TwinsUK laboratories within 24 h of collection (stool) and 2 h of collection (saliva), where they were stored at  $-80^{\circ}\text{C}$  until processing. Genomic DNA was extracted and processed at Clinical Microbiomics (Copenhagen, Denmark) for shotgun metagenomics using 2  $\times$  150 bp paired-end chemistry on an Illumina NovaSeq 6000 platform. Quality control of sequencing data was performed by YAMP [16]. Taxonomic profiling was carried out using kracken2/bracken [17], and HUMAnN3 [18] was used for functional profiling (UniRef90 reference database).

### Microbiome profiling: NOAR cohort

Patients with chronic RA were obtained from the Norfolk Arthritis Register (NOAR) (Participant Characteristics [Supplementary Table 1](#), available at *Rheumatology* online). NOAR samples (212 stool and saliva samples) were collected from participants on the registry with RA who had been

taking DMARD therapy >12 months [19], with all participants completing the clinical questionnaire. Samples were sent to Clinical Microbiomics (Denmark) and processed through default settings in YAMP, as outlined above.

## Statistical analysis

### Alpha- and beta-diversity

Chao1 richness, Shannon- and Simpson diversity index were calculated in the Phyloseq R package (v1.40.0) [20]. Alpha-diversity metrics were compared at each time point using linear mixed modelling, adjusting for clinical and demographic covariates such as age, sex, body mass index (BMI), smoking status, ethnicity and hospital centre, within the Vegan R package (v2.6–4) [21].

Before beta-diversity, data were filtered using PERfect (v1.10.0) [22], a novel permutation filter approach based on statistical hypothesis testing in taxa reduction. PERfect removes potential sequencing contaminant and taxa with minimal impact on the overall covariation within microbiome datasets.

Data were normalized with an arcsine square root transformation and clusters were visualized using Bray–Curtis dissimilarity and principal coordinates analysis (PCoA) using Vegan. Permutational Multivariate Analysis of Variation (PERMANOVA) was used to statistically compare cluster centroids and repeat measures were considered using *strata* parameter, where applicable.

### Differential abundance

Comparison of different taxonomic abundances between time points and clinical data was carried out using the limma:voom pipeline (v3.52.4) [23, 24] after normalizing abundances using GMPR (v0.1.3) [25]. Statistical significance threshold was set at 10% false discovery rate (FDR). Clinical Disease Activity Index (CDAI) score was used to classify patients into two groups according to minimal clinically important improvement (MCII), as defined in Gupta and colleagues [13], to determine predictive DMARD treatment response in baseline microbiota. This allowed the comparison of those with a meaningful clinical response to DMARD therapy with participants that did not (Supplementary Methods, available at *Rheumatology* online).

### Predictive model for MCII

A sparse partial least squares discriminant analysis [(s)PLS-DA] model was developed to predict MCII in microbiome data, using the MixOmics package (v6.20.0) [26]. Baseline abundance and MCII status established after six weeks of DMARD treatment were used to train and test the model. Samples were randomly split into two equal subsets, between those that did and did not achieve MCII ( $n = 53$ , with  $n = 32$  MCII+ and  $n = 21$  MCII-), and confusion matrices and area under the curve (AUC) plots were used to evaluate model performance. An additional microbiome dataset obtained from Gupta and colleagues [13] was used as a validation RA cohort for the (s)PLS-DA model.

## Results

### Sample description

The IMRABIOME cohort comprised 144 RA participants assessed at baseline ( $n = 144$ ), at six weeks ( $n = 117$ ) and at 12 weeks ( $n = 95$ ) follow-up. A total of 72.9% of participants

were female and 63.8% northern Europeans. Stool and saliva samples were collected at all timepoints for metagenomics alongside clinical markers, demographic and lifestyle information. All data were available at all three timepoints for 92 participants, 27 had data available at two time points (i.e. baseline and six weeks, or 12 weeks); and 26 with data available only at a single time point (i.e. only baseline).

Demographics and clinical characteristics are presented in Table 1.

### Alpha diversity

Patient demographics impact on alpha-diversity was examined using Chao1 richness, Shannon- and Simpson's diversity index, in stool and saliva. No significant difference was seen in alpha-diversity metrics at any timepoint for either sample type ( $P > 0.05$ ). Smoking status was significant negatively associated with Shannon and Simpson indices across all three timepoints in saliva (Supplementary Table 2, available at *Rheumatology* online), with current and previous smokers having a reduced diversity compared with non-smokers. Smoking status did not significantly associate with alpha-diversity in the stool microbiome.

Disease activity measured by Disease Activity Score (DAS)28 and CDAI was negatively associated with all three alpha-diversity metrics in stool (Table 2). Sulfasalazine reduced richness in the gut at six weeks ( $P = 0.015$ ), no further DMARD influence was seen in gut or saliva alpha-diversity. Other demographics, such as sex, associated with Shannon diversity at baseline and six weeks, with males showing a reduced diversity compared with females ( $P < 0.05$ ).

### Beta diversity

Microbial data were analysed with arcsine transformed PERfect-filtered data using PCoA to identify outliers before statistical comparisons. After removing outliers, we recalculated Bray–Curtis dissimilarity and used PERMANOVA models to establish demographic and clinical factors contributing to gut and saliva microbiome variance. Using univariate and multivariate models we assessed impacts of age, sex, BMI, smoking, hospital centre, ethnicity, disease severity as measured by DAS28 score, and DMARD use, at six and 12 weeks.

In multivariate models, the saliva microbiome was significantly different in participant smoking status at all timepoints ( $P < 0.001$ , Supplementary Table 3, available at *Rheumatology* online). Age was the only other variable to significantly differ in the saliva microbiome, which occurred at baseline ( $P = 0.016$ ). In the gut, age and ethnicity significantly differed at all timepoints, while BMI and DAS28 also differed at individual sampling points (Table 3). DMARD monotherapy did not significantly impact microbial communities in stool or saliva models; however, at six weeks' follow-up, for participants on combination therapy, the methotrexate component had a significant effect on gut microbiome beta-diversity.

### Differential abundance

Differential abundance (DA) was carried out using the limma:voom approach with abundance normalized with the GMPR, characterized by improved power and control of false positives compared with other normalization techniques. First, we checked for microbiome abundance changes across time points using linear contrast for pairwise comparisons

**Table 1.** Clinical and demographic characteristics of RA patients

Trait		Baseline, <i>n</i> = 144	6 weeks, <i>n</i> = 117	12 weeks, <i>n</i> = 95	
Sex <sup>a</sup>	Female, <i>n</i> (%)	105 (72.4)			
	Male, <i>n</i> (%)	38 (26.2)			
	Unknown, <i>n</i> (%)	2 (1.4)			
Age, mean ±SD		53.8 ± 13.9			
Smoking status <sup>a</sup>	Never, <i>n</i> (%)	66 (45.5)			
	Ever, <i>n</i> (%)	52 (35.9)			
	Current, <i>n</i> (%)	24 (16.6)			
	Unknown, <i>n</i> (%)	3 (2.1)			
Diet	Diet including meat and fish, <i>n</i> (%)	65 (49.3)			
	Diet excluding meat and fish, <i>n</i> (%)	67 (50.7)			
Body mass index, mean ±SD		28.5 ± 6.1			
Ethnicity	Northern European, <i>n</i> (%)	92 (63.8)			
	South Asian, <i>n</i> (%)	17 (11.8)			
	East Asian, <i>n</i> (%)	5 (3.4)			
	Black African, <i>n</i> (%)	9 (6.2)			
	Black Caribbean, <i>n</i> (%)	7 (4.8)			
	Mixed – Other, <i>n</i> (%)	4 (2.7)			
	Unknown, <i>n</i> (%)	10 (6.9)			
Hospital Centre	North, <i>n</i> (%)	13 (9.0)			
	South, <i>n</i> (%)	91 (63.1)			
	East, <i>n</i> (%)	19 (13.1)			
	West, <i>n</i> (%)	21 (14.5)			
Anti-CCP	Positive, <i>n</i> (%)	88 (65.7)			
	Negative, <i>n</i> (%)	46 (34.3)			
Tender joint score	Median (min, max)	7 (0; 28)	3 (0; 27)	2 (0; 28)	
Swollen joint score	Median (min, max)	5 (0; 28)	1 (0; 22)	1 (0; 23)	
C-reactive protein	Median (min, max)	7.1 (1; 209)	4 (1; 78)	4 (1; 122)	
DAS28 score	Median (min, max)	4.9 (1.4; 8.2)	3.5 (0.0; 7.3)	3.8 (0.7; 7.7)	
	High, <i>n</i> (%)	65 (45.1)	22 (18.8)	19 (20.0)	
	Moderate, <i>n</i> (%)	54 (37.5)	42 (35.9)	21 (22.1)	
	Low, <i>n</i> (%)	22 (15.3)	50 (42.7)	51 (53.7)	
	Unknown, <i>n</i> (%)	3 (2.1)	3 (2.6)	4 (4.2)	
	Disease activity decreased <i>vs</i> baseline <sup>b</sup>			54/112 (48.2)	44/88 (50.0)
	CDAI score	Median (min, max)	25 (0; 73)	13 (0; 65)	8.5 (0; 61)
	High, <i>n</i> (%)	80 (55.6)	30 (25.6)	21 (22.1)	
	Moderate, <i>n</i> (%)	33 (22.9)	34 (29.1)	17 (17.9)	
	Low, <i>n</i> (%)	23 (16.0)	32 (27.4)	26 (27.4)	
	Remission/no active disease, <i>n</i> (%)	4 (2.8)	17 (14.5)	26 (27.4)	
	Unknown, <i>n</i> (%)	4 (2.8)	4 (3.4)	5 (5.3)	
	Disease activity decreased <i>vs</i> baseline <sup>c</sup>		64/106 (60.4)	49/82 (59.8)	
DMARD Treatment					
	Monotherapy	Methotrexate – monotherapy, <i>n</i>	75	54	32
		Sulfasalazine – monotherapy, <i>n</i>	4	3	3
Hydroxychloroquine – monotherapy, <i>n</i>		4	4	3	
Combination	Methotrexate and Sulfasalazine, <i>n</i>	7	12	6	
	Methotrexate and Hydroxychloroquine, <i>n</i>	51	37	37	
	Sulfasalazine and Hydroxychloroquine, <i>n</i>	2	2	3	
	Methotrexate, Sulfasalazine and Hydroxychloroquine, <i>n</i>	1	2	2	
Total use (any)	Methotrexate, <i>n</i> (%)	134/144 (93.1)	105/113 (92.9)	77/86 (89.5)	
	Sulfasalazine, <i>n</i> (%)	16/144 (11.1)	19/109 (17.4)	14/85 (16.5)	
	Hydroxychloroquine, <i>n</i> (%)	58/144 (40.3)	45/110 (40.9)	45/87 (51.7)	

<sup>a</sup> One patient did not have microbiome data at baseline, and at follow-up only; accordingly, proportions for sex and smoking at baseline are estimated for 145 individuals.

<sup>b</sup> Patients' disease activity category decreased from high to moderate or low, from moderate to low.

<sup>c</sup> CDAI score decreased for at least 1 unit for those who had low activity, at least 6 units for those who had moderate activity, and at least 12 units for those who had high activity.

RA: rheumatoid arthritis; BMI: body mass index; DAS: Disease Activity Score.

(Supplementary Tables 4 and 5, available at *Rheumatology* online). Differences in the saliva microbiome were inconsistent, as such we focused on the gut microbiome (saliva data in Supplementary Tables 6 and 7, available at *Rheumatology* online). In the gut, when comparing timepoints, we found more differential abundant taxa between baseline and 12 weeks, than baseline and six weeks (Supplementary Table

8, available at *Rheumatology* online). No significant findings were discovered while comparing follow-up time points. Next, we compared our study cohort of early-RA patients to NOAR, a long-term treatment cohort. DMARD-naive baseline, several weeks of DMARD treatment (i.e. combination of six-week and 12-week samples) and long-term (>1 year) use of DMARD (NOAR sample) were compared (FDR < 0.10)

**Table 2.** Disease activity and microbiome alpha-diversity

Alpha diversity measure	Disease activity and severity scores	Baseline	6 weeks	12 weeks
Chao1 richness	DAS28	$\beta = -161.64 \pm 42.55; P = 0.0002$	$\beta = -142.15 \pm 38.97; P = 0.0004$	$\beta = -197.98 \pm 55.06; P = 0.0005$
	CDAI	$\beta = -16.03 \pm 4.12; P = 0.0001$	$\beta = -15.38 \pm 4.48; P = 0.0008$	$\beta = -19.92 \pm 6.10; P = 0.001$
Shannon index	DAS28	$\beta = -0.59 \pm 0.19; P = 0.003$	$\beta = -0.46 \pm 0.21; P = 0.027$	$\beta = -0.49 \pm 0.29; P = 0.094$
	CDAI	$\beta = -5.51 \pm 2.07; P = 0.009$	$\beta = -3.46 \pm 1.83; P = 0.062$	$\beta = -5.14 \pm 2.69; P = 0.059$
Simpson index	DAS28	$\beta = -2.20 \pm 1.11; P = 0.050$	$\beta = -3.19 \pm 1.11; P = 0.005$	$\beta = -2.01 \pm 1.59; P = 0.211$
	CDAI	$\beta = -19.10 \pm 12.03; P = 0.114$	$\beta = -29.56 \pm 9.76; P = 0.003$	$\beta = -24.75 \pm 14.11; P = 0.095$

Alpha diversity was calculated from stool metagenomes. Estimates  $\pm$  standard error and *P*-values for multiple regression of RA activity scores on Chao1, Shannon and Simpson indices; adjusted for age, sex, BMI and smoking status.

BMI: body mass index; CDAI: Clinical Disease Activity Index; DAS: Disease Activity Score; RA: rheumatoid arthritis.

**Table 3.** Disease activity, patient characteristics and gut microbiome variance

Trait	Model	Baseline		Six weeks		12 weeks	
		R <sup>2</sup>	<i>P</i> -value	R <sup>2</sup>	<i>P</i> -value	R <sup>2</sup>	<i>P</i> -value
Age	Univariate	1.4	0.0186	1.8	0.0052	2.5	0.0031
	Adjusted	1.3	0.0288	2.0	0.0091	3.6	0.0002
Sex	Univariate	1.3	0.0300	1.6	0.0205	1.6	0.0617
	Adjusted	1.0	0.1091	1.8	0.0222	1.5	0.1591
BMI	Univariate	1.6	0.0079	0.9	0.2834	1.0	0.5027
	Adjusted	1.4	0.0193	1.1	0.2833	0.9	0.6611
Smoking	Univariate	1.0	0.8588	1.4	0.8074	1.7	0.7808
	Adjusted	1.2	0.7127	1.6	0.7723	2.4	0.4017
Diet	Univariate	0.3	0.9529	—	—	—	—
	Adjusted	0.4	0.9738	—	—	—	—
Site (hospital)	Univariate	2.2	0.3239	3.1	0.1496	4.1	0.0810
	Adjusted	2.3	0.3476	3.1	0.3599	4.9	0.0505
Ethnicity	Univariate	7.1	0.0003	7.6	0.0011	8.8	0.0033
	Adjusted	6.8	0.0002	7.3	0.0083	8.4	0.0262
DAS28 score	Univariate	1.4	0.0173	1.6	0.0257	3.4	0.0004
	Adjusted	0.9	0.1824	0.8	0.5704	2.3	0.0154
Tender joint score	Univariate	0.9	0.1605	—	—	—	—
	Adjusted	1.1	0.0649	—	—	—	—
Swollen joint score	Univariate	0.5	0.6861	—	—	—	—
	Adjusted	1.1	0.0728	—	—	—	—
C-reactive protein	Univariate	0.5	0.5949	—	—	—	—
	Adjusted	0.6	0.5206	—	—	—	—
Methotrexate–mono	Univariate	—	—	3.4	0.0139	3.2	0.2030
	Adjusted	—	—	2.2	0.2237	2.4	0.4633
Sulfasalazine–mono	Univariate	—	—	1.0	0.2898	1.2	0.4938
	Adjusted	—	—	—	—	—	—
Hydroxychloroquine–mono	Univariate	—	—	1.2	0.8384	2.1	0.5215
	Adjusted	—	—	—	—	—	—
Use of methotrexate	Univariate	—	—	1.6	0.0287	3.4	0.3447
	Adjusted	—	—	1.8	0.0353	1.4	0.2626
Use of sulfasalazine	Univariate	—	—	1.6	0.0283	2.1	0.0211
	Adjusted	—	—	1.1	0.2645	1.5	0.1735
Use of hydroxychloroquine	Univariate	—	—	1.5	0.0452	1.1	0.3945
	Adjusted	—	—	0.8	0.6492	1.2	0.3827

Bray–Curtis dissimilarity was calculated using stool microbiome at species level; because of high correlation between DAS28 and CDAI scores, we chose to analyse DAS28 score only.

BMI: body mass index; CDAI: Clinical Disease Activity Index; DAS28: Disease Activity Score 28; mono: monotherapy.

(Supplementary Tables 9–11, available at *Rheumatology* online). In total, 32 taxa differed in abundance between long-term DMARD use and several weeks of DMARD use. In contrast, 86 taxa significantly decreased in long-term DMARD treated participants compared with DMARD-naïve baseline, with *Porphyromonas gingivalis* and 15 *Prevotella* spp. among the reduced taxa.

Minimal clinically important improvement (MCII) was used to assess microbiome change based on patient response

to DMARD treatment. At baseline, there were no detectable microbiome differences between those that achieved MCII and those who did not. After six weeks of DMARD treatment, 18 taxa differed between MCII groups, of which 13 decreased and five increased in abundance, in those that achieved MCII vs those that did not. Of the 13 decreased taxa, seven (54%) were *Prevotella* spp. At 12 weeks treatment, a further 13 taxa significantly decreased when comparing participants that fulfilled MCII compared with those that

did not, of which 10 were *Streptococcus* spp. (Table 5). Overall, these findings led us to conclude that the gut microbiome normalises in RA patients with longer treatment, but only in patients who respond to DMARD treatment.

### Anti CCP

Anti-cyclic citrullinated peptide (anti-CCP) values at baseline were positive for 88 participants and negative for 46 participants (defined as <20 u/ml antibody negative,  $\geq 20$  u/ml positive), Table 1. We explored the relationship between anti-CCP antibodies and smoking in saliva microbiota using linear regression models. The inclusion of anti-CCP did not significantly alter Shannon diversity ( $P=0.11$ ) or Chao1 richness ( $P=0.67$ ) measures. Furthermore, univariate ( $P=0.22$ ) and multivariate ( $P=0.20$ ) PERMANOVA statistics suggested anti-CCP did not significantly impact the saliva microbiota diversity. The inclusion of anti-CCP in the gut microbiota linear regression did not improve model outcomes, with no significant association between anti-CCP and the gut microbiota (Shannon  $P=0.15$ , Chao1  $P=0.07$ ). *Prevotella* species were, however, elevated in individuals positive for anti-CCP antibodies (Supplementary Table 12, available at *Rheumatology* online).

### Predictive model for MCII

We applied sPLSDA analysis to develop a model to predict MCII status upon short-term treatment (see 'Materials and methods'). We used baseline species abundances and MCII status as established at six weeks. Furthermore, using Kegg Orthology and Metacyc database microbial gene and biochemical pathways, predictive potential was also explored. To generate train and test samples, we randomly split the dataset with available information into two equal subsets,  $n=53$  each, with  $n=32$  who achieved MCII and  $n=21$  who did not. In the gut, the final model for microbes provided AUC=0.66 ( $P=0.0539$ ); while for Gupta and colleagues [13], AUC=0.60 ( $P=0.3706$ ) (Fig. 1). Saliva samples manifested an AUC=0.61 ( $P=0.14$ ) for microbes and AUC=0.69 ( $P=0.033$ ) for Metacyc pathways, while KO genes for stool and saliva had AUCs <0.55 (Supplementary Tables 13–15, available at *Rheumatology* online).

### Discussion

The microbiome has an established and prominent role in RA disease and early-onset development [27, 28]. Our research focused on early-disease RA patients who were DMARD-naive and examined microbiome changes during initiation of DMARD therapy in a large-scale observation study of oral and gut microbiome in 144 individuals. DMARD therapy is recommended early in disease to improve patient outcomes, but currently a trial-and-error approach is used, with drug dose escalation if response is poor. The ability to determine who will respond to DMARDs would enable a personalised approach to treatment selection, rather than the ubiquitous administration of methotrexate in new-onset RA in the UK. Our interest was in being able to predict how patients will respond to DMARD therapy. There were two main findings from our study. First, microbial abundance among patients that attained MCII (MCII+) and those that did not (MCII-), were different at 6 and 12 weeks of DMARD treatment. In total, 26 different stool microbes decreased in MCII+ compared with MCII-. *Prevotella* spp. and *Streptococcus* spp. were the

predominant taxa to decline after six weeks and 12 weeks, respectively. Interestingly, diet preference had limited impact on the stool microbiota and abundance differences seen in this study, with no changes to microbiota seen (Tables 4 and 5). Second, we found the DMARD-naive microbiome was indicative of clinical response. Metacyc pathways best predicted MCII+ from DMARD-naive samples, particularly saliva, providing a promising target for metabolomics. Furthermore, we found 86 microbes decreased significantly in long-term DMARD-exposed participants from the NOAR cohort, when compared with DMARD-naive patients; *Prevotella* spp. accounted for ~20% of species in decline.

*Prevotella* are commonly associated with RA pathogenesis. In humans, these commensal microbes colonise the skin, vagina, oral cavity, respiratory tract and intestine, and have important roles in metabolism and health [29]. Here we add further evidence in support of gut *Prevotella* spp. involvement in RA prognosis, by contrasting MCII+ and MCII-. After six weeks of DMARD treatment, 18 microbes significantly differed in patients' gut microbiome showing clinical response to therapy, of those 13 decreased including seven *Prevotella* spp. (see Table 4). Furthermore, 15 *Prevotella* spp. were identified to decline within long-term DMARD-treated NOAR participants, when compared with early-onset DMARD-naive patients. In total, 15 different *Prevotella* spp. were negatively associated with clinical improvement to therapy, including *P. denialis*, *P. denticola*, *P. histicola*, *P. intermedia*, *P. melaninogenica*, *P. nigrescens*, *P. oris* and *P. ruminicola*. These species have been linked to respiratory disease, multiple sclerosis, brain abscess, blood and cardiac infection, musculoskeletal infection, and most commonly periodontal disease [29]. Previously, we reported *Prevotella* spp. play a role in RA development, and microbial abundance increased in those with the genotype predisposing risk for RA even in the absence of disease [5]. Interestingly, we did not detect significant associations between *P. copri*, DMARD treatment and patient clinical response. It is well established that *P. copri* is involved in early RA, with increased abundance thought to stimulate expression of anti-citrullinated proteins and serum antibodies [6, 8]. Genetic factors predispose to RA, with additive influence of acquired environmental risk [30]. Host genetic factors mediate the microbiome and are thought to lead to a pro-inflammatory, hyper-active immune type [5, 30]. The presence of anti-CCP antibodies at baseline may be considered to represent the shared epitope risk, in line with our previous study demonstrating increased *Prevotella* spp. in those positive for the shared epitope [5]. Regression models for stool including anti-CCP antibody status as covariate showed no influence on alpha-diversity in our sample. We identified seven *Prevotella* spp. to be significantly elevated in those having anti-CCP antibodies compared with those without, and several other species including *P. copri* showed a similar trend (Supplementary Table 12, available at *Rheumatology* online) [31].

Of metagenomics studies investigating RA and the microbiome; two reported taxa differences between participants who respond well to DMARD therapies compared with those who do not [12, 13]. In line with our findings, no differential abundance in *P. copri* was reported by either group. Whilst a common theme across RA literature, we speculate that abundance of *P. copri* is more prominent when comparing healthy and disease participants, which may explain the lack of association in this study. *P. copri* comprises four genetically

distinct clades with genetic and functional differences that are influenced by diet and lifestyle [32–34]. For example, in plant-rich ‘non-westernised’ diets, *P. copri* can catabolise plant carbohydrates and fibres; alternatively, in higher protein diets *P. copri* is known to break down branch chain amino acids (BCAAs) from meats [29, 32, 34]. Interestingly, in a recent study gut isolates of *P. copri* strains were shown to differ in RA participants, with more severe arthritis induced in mice from RA patient strains than from healthy controls [34]. These findings warrant further investigation and may extend more broadly to *Prevotella* spp. Of note, periodontal disease appears to commonly be associated with the *Prevotella* spp. we identified (e.g. *P. denalis*, *P. denticola*, *P. histicola*, *P.*

*intermedia*, *P. melaninogenica*, *P. nigrescens*, *P. oris*, *P. ruminicola*) with poor treatment outcome [29].

Periodontal disease is more frequent in RA [35]. Despite environmental niches and segregation, it is common for microbes to translocate from the oral cavity to the gut. Common examples in RA are *Porphyromonas gingivalis* and *P. copri*; however, *Streptococcus* spp. are also associated with oral to gut translocation [36]. In the oral cavity, streptococci are known to form early-plaque biofilms that can support microbes associated with gingival disease, such as *P. gingivalis* [37]. Intriguingly, *P. gingivalis* was identified in higher abundance in baseline samples when compared with long-term DMARD participants in NOAR. Furthermore *Streptococcus* spp. significantly decreased in the gut of early RA patients having MCII+ after 12 weeks of DMARD treatment, in this study.

Our findings support the hypothesis of DMARD restoration of a eubiotic gut microbiome when patient and treatment

**Table 4.** Patient response MCII+/MCII- and differentially abundant taxa during the first six weeks of DMARD treatment

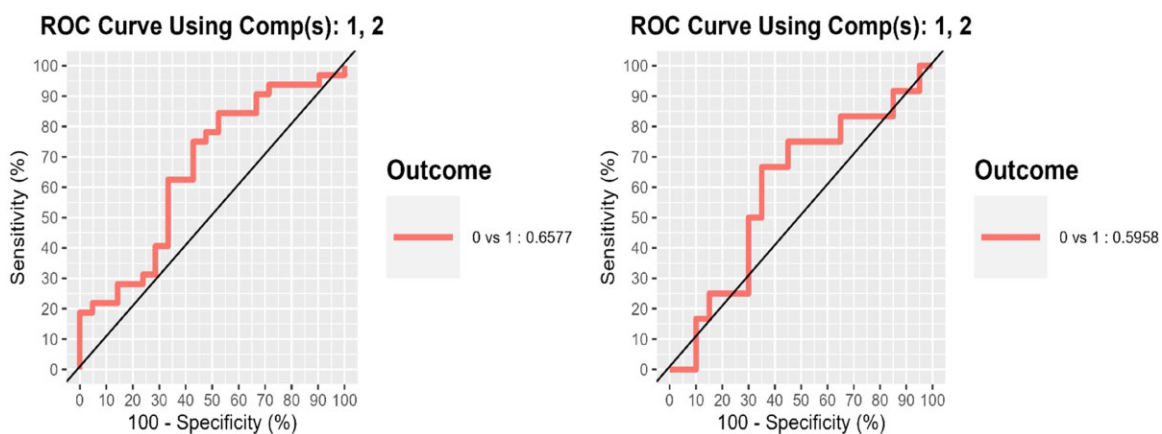
Taxa	MCII+ vs MCII- @ Six weeks		
	logFC	P-value	adj. P-value
<i>Phocaeicola salanitronis</i>	-1.73	0.000	0.002
<i>Bacteroides coprosuis</i>	-1.68	0.000	0.002
<i>Phocaeicola coprophilus</i>	-1.87	0.000	0.008
Candidatus Arthromitus sp. SFB rat Yit	0.84	0.000	0.039
<i>Prevotella jejuni</i>	-1.31	0.000	0.039
<i>Prevotella intermedia</i>	-1.35	0.000	0.047
<i>Prevotella nigrescens</i>	-1.38	0.000	0.047
<i>Propionibacterium australiense</i>	0.74	0.001	0.078
<i>Prevotella fusca</i>	-1.28	0.001	0.078
<i>Prevotella oris</i>	-1.31	0.001	0.078
<i>Streptococcus sobrinus</i>	-1.39	0.001	0.078
<i>Paraprevotella xyliniphila</i>	-1.98	0.001	0.078
<i>Prevotella</i> sp. WR041	-1.31	0.001	0.078
<i>Rhodopseudomonas palustris</i>	0.65	0.001	0.078
<i>Paludibacterium</i> sp. B53371	0.75	0.001	0.078
<i>Nitrogeniibacter mangrovi</i>	0.76	0.001	0.078
<i>Prevotella enoeca</i>	-1.19	0.002	0.094
<i>Parabacteroides distasonis</i>	-1.49	0.002	0.094

Differentially abundant taxa between RA patients who achieved minimum clinically important improvement (MCII+) and who did not (MCII-) during the first six weeks of DMARD treatment. Linear contrasts have been constructed and tested for significance using voom: limma approach combined with GMPR normalisation. Adjustment for multiple testing was done using Benjamini–Hochberg false discovery rate within each taxonomic level. Reported are findings with FDR <10% in at least one contrast. No significant results were obtained for comparisons between MCII+ and MCII- at baseline.

**Table 5.** Patient response MCII+/MCII- and differentially abundant taxa following 12 weeks of DMARD treatment

Taxa	MCII+ vs MCII- @ 12 weeks		
	logFC	P-value	adj. P-value
<i>Gemella haemolysans</i>	-1.7	0.000	0.094
<i>Staphylococcus epidermidis</i>	-1.45	0.000	0.094
<i>Veillonella parvula</i>	-1.92	0.000	0.094
<i>Streptococcus mitis</i>	-1.76	0.000	0.094
<i>Streptococcus</i> sp. oral taxon 061	-1.75	0.001	0.094
<i>Streptococcus gordonii</i>	-1.61	0.001	0.094
<i>Streptococcus oralis</i>	-1.59	0.001	0.094
<i>Streptococcus</i> sp. A12	-1.85	0.001	0.094
<i>Streptococcus lactarius</i>	-1.78	0.001	0.094
<i>Streptococcus gallolyticus</i>	-1.52	0.001	0.094
<i>Streptococcus pneumoniae</i>	-1.21	0.001	0.094
<i>Streptococcus parasanguinis</i>	-1.89	0.001	0.094
<i>Streptococcus</i> sp. LPB0220	-2.02	0.001	0.094

Differentially abundant taxa between RA patients who achieved minimum clinically important improvement (MCII+) and who did not (MCII-) following 12 weeks of DMARD treatment. Linear contrasts have been constructed and tested for significance using voom: limma approach combined with GMPR normalisation. Adjustment for multiple testing was done using Benjamini–Hochberg false discovery rate within each taxonomic level. Reported are findings with FDR <10% in at least one contrast. No significant results were obtained for comparisons between MCII+ and MCII- at baseline.



**Figure 1.** Area under the ROC curve for predictive model of MCII after short-term treatment. Left panel is for the test sample resulted from the split of our dataset into random train and test samples. Right panel is for external data [13]

align [38, 39]. We had anticipated finding baseline microbiome samples predictive of response to treatment. While baseline differential abundance analysis did not discriminate responders (MCII+) from non-responders (MCII-), longitudinal analysis showed changing microbiota and positive response to DMARDs. At 6 weeks there was a reduction of multiple species of *Prevotella* in responsive participants. At 12 weeks, multiple species of Streptococci were reduced in responders. Methotrexate was associated with beta-diversity difference following six weeks of treatment, even when adjusting for disease activity and other confounders (age, sex, BMI, smoking status, ethnicity, hospital centre, diet and other DMARDs).

Predictive tools for RA treatment outcome utilising oral and gut microbiome are advancing [12–14]. In a small Hispanic cohort ( $n=26$ ) a model using DMARD-naive gut metagenomes from patients receiving monotherapy methotrexate was able to predict non-responsiveness [12]. Gupta and colleagues [13] retrospectively used gut metagenomes from patients ( $n=32$ ) taking DMARDs to predict MCII response with 90% accuracy using a neural network. Similarly, Zhang and colleagues [14] accurately predicted DAS28-ESR response in using long-term DMARD-treated patients and controls, with dental metagenome samples. Our models used gut and oral metagenomes to predict DMARD-naive MCII response. Although our models were not as strong as those cited above, we were able to better identify MCII+ as highlighted by our sensitivity scores (Supplementary Tables 13–15, available at *Rheumatology* online). Northern Europeans accounted for >60% of our sample and patients often received combination therapy to treat RA. The real strength of our work lies in our validation using two external data cohorts, a novel and key contrast to similar studies that may have overfit models with small sample sizes and a lack of external validation. Another strength is that the population cohort of DMARD-naive participants permits the exploration of the influence of therapy.

The range of DAS28 values at baseline was large, although the lower quantile cut-off at baseline was 3.6, therefore comprising those with clinically meaningful disease activity. When low DAS scores were removed from the analysis, the negative linear relationship with alpha diversity was maintained. This association decreased with received treatment; for example, DAS28 negatively associated with alpha-diversity at baseline, although an improvement in DAS28 is seen in patients at six (48.2%) and 12 weeks (50%) (Table 1) following DMARD treatment. Interestingly, when broken down the tender joint score was the principal DAS28 measure that negatively associated with alpha-diversity.

Long-term DMARD comparison came from NOAR: study participants registered to assist arthritis research. Disease activity was from medical records and was participant generated, not clinician scored, so we used available data to calculate DAS28-CRP. Differential abundance analysis in NOAR using the criteria cut-point of DAS28-CRP <2.6 showed one taxa altered, *Actinomyces sp. Oral taxon 414*, between high- vs low disease activity (Supplementary Table 11, available at *Rheumatology* online). Furthermore, the predictive model used in the drug-naive cohort resulted in similar AUC for gut microbiota at 0.65 AUC (Supplementary Table 13, available at *Rheumatology* online).

We acknowledge that there are limitations to this study. Changes in saliva metagenomes were sparse, with smoking

status and the predictive model with metacyc pathways being the only clear signals achieved. Similar predictive results were seen in Zhang and colleagues when examining saliva; however, dental microbiota were able to predict response [14]. The longitudinal data collection was for 12 weeks post recruitment, though longer follow-up of 6–12 months would have been preferable.

A strength of our work is the ethnic mix recruited to the study reflecting an inner-city UK sample: we found beta-diversity to be reduced in northern Europeans compared with south Asians at baseline, likely driven by the small sub-group numbers, but there was no impact of ethnicity on predictive models. Interestingly, *Prevotella* spp. are known to dominate the gut microbiome in ‘non-westernised’ populations [29]. The interplay between ethnicity, the microbiome and RA warrants investigation and may give further insight into patient-specific treatment.

In summary, we identified a partial restoration of the microbiome to a more eubiotic state in RA patients at 6 weeks and 12 weeks DMARD treatment in participants that responded well to DMARD therapy. This was further supported by long-term (>1 year) treated DMARD RA participants with similar community shifts. Finally, microbiomes provide a promising diagnostic tool for guiding therapeutic decisions in future.

## Supplementary material

Supplementary material is available at *Rheumatology* online.

## Data availability

The data underlying this article are available in Sequence Read Archive (SRA) at <https://dataview.ncbi.nlm.nih.gov/object/PRJNA957107?reviewer=e4pj6rso2m8osj2c6hilgqn9ei>, and can be accessed with accession number PRJNA957107.

## Author contributions

Nathan P. Danckert (Data analysis, Methodology, Writing, Writing—Review & Editing), Maxim B. Freidin (Study design, Data analysis, Methodology, Writing, Writing—Review & Editing, Funding acquisition), Isabelle Granville Smith (Data curation, Writing, Writing—Review & Editing), Philippa M. Wells (Data analysis, Methodology, Writing), Maryam Kazemi Naeini (Methodology, Writing—Review & Editing), Alessia Visconti (Data analysis, Methodology), Roger Compte (Methodology, Writing—Review & Editing), Alexander MacGregor (Writing—Review & Editing, Funding acquisition), Frances M. K. Williams (Study design, Data curation, Data analysis, Methodology, Writing, Writing—Review & Editing, Funding acquisition). All authors read and approved the final manuscript.

## Funding

This work has been supported by funding from Versus Arthritis (grant number 21227).

**Disclosure statement:** The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest to declare that are relevant to the content of this article. All authors certify that they have no affiliations



with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

## Acknowledgements

The authors would like to thank the TwinsUK laboratory team, the participants of the Imrabiome and NOAR cohorts, and the hospital centres that participated in this study: Guy's Rheumatology Department; Guy's and St Thomas' NHS Foundation Trust, with particular thanks to Quincy D'Souza, Profs Andrew Cope and Bruce Kirkham; the research staff and nurses at the Rheumatology Departments in: King's College Hospital (KCH) – King's College Hospital NHS Foundation Trust; University Hospital Lewisham – Lewisham and Greenwich NHS Foundation Trust; Ashford Hospital – Ashford and St Peter's Hospitals NHS Foundation; Kingston Hospital – Kingston Hospital NHS Foundation Trust; Addenbrooke's Hospital – Cambridge University Hospitals NHS Foundation Trust NHS; Lister Hospital – East and North Hertfordshire NHS Trust; Southend University Hospital – Southend University Hospital NHS; Basildon University Hospital – Basildon and Thurrock University Hospitals NHS Foundation Trust; Barts Hospital – Barts Health NHS Trust; Croydon Health Services – Croydon Health Services NHS Trust; Norfolk Arthritis Registry (NOAR) and University of East Anglia (UEA) with special thanks to Karen Durant and Jack Dainty. We thank Prof. Julian Marchesi for his professional opinion on the microbiota.

## References

- MacGregor AJ, Snieder H, Rigby AS *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000;43:30–7.
- Daikh DI. Rheumatoid arthritis: evolving recognition of a common disease. *Best Pract Res Clin Rheumatol* 2022;36:101740.
- Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell* 2012;148:1258–70.
- Gilbert JA, Blaser MJ, Caporaso JG *et al.* Current understanding of the human microbiome. *Nat Med* 2018;24:392–400.
- Wells PM, Adebayo AS, Bowyer RCE *et al.* Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study. *Lancet Rheumatol* 2020;2:e418–27.
- Scher JU, Sczesnak A, Longman RS *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *Elife* 2013;2:e01202.
- Alghamdi MA, Redwan EM. Interplay of microbiota and citrullination in the immunopathogenesis of rheumatoid arthritis. *Probiotics Antimicrob Proteins* 2022;14:99–113.
- Pianta A, Arvikar S, Strle K *et al.* Evidence of the immune relevance of *Prevotella copri*, a gut microbe, in patients with rheumatoid arthritis. *Arthritis Rheumatol* 2017;69:964–75.
- Cronstein BN, Aune TM. Methotrexate and its mechanisms of action in inflammatory arthritis. *Nat Rev Rheumatol* 2020;16:145–54.
- Scher JU, Nayak RR, Ubeda C, Turnbaugh PJ, Abramson SB. Pharmacomicrobiomics in inflammatory arthritis: gut microbiome as modulator of therapeutic response. *Nat Rev Rheumatol* 2020;16:282–92.
- Emery P, Breedveld FC, Hall S *et al.* Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial. *Lancet* 2008;372:375–82.
- Artacho A, Isaac S, Nayak R *et al.* The pretreatment gut microbiome is associated with lack of response to methotrexate in new-onset rheumatoid arthritis. *Arthritis Rheumatol* 2021;73:931–42.
- Gupta VK, Cunningham KY, Hur B *et al.* Gut microbial determinants of clinically important improvement in patients with rheumatoid arthritis. *Genome Med* 2021;13:149.
- Zhang X, Zhang D, Jia H *et al.* The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat Med* 2015;21:895–905.
- Aletaha D, Neogi T, Silman AJ *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* 2010;62:2569–81.
- Visconti A, Martin TC, Falchi M. YAMP: a containerized workflow enabling reproducibility in metagenomics research. *GigaScience* 2018;7:1–9.
- Lu J, Rincon N, Wood DE *et al.* Metagenome analysis using the Kraken software suite. *Nature Protocols* 2022;17:2815–39.
- Beghini F, McIver LJ, Blanco-Miguez A *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 2021;10:e65088.
- Symmons D, Silman A. The Norfolk arthritis register (NOAR). *Clin Exp Rheumatol* 2003;21:S94–9.
- McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 2013;8:e61217.
- Dixon P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* 2003;14:927–30.
- Smirnova E, Huzurbazar S, Jafari F. PERFect: PERmutation Filtering test for microbiome data. *Biostatistics* 2019;20:615–31.
- Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014;15:R29.
- Ritchie ME, Phipson B, Wu D *et al.* limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res* 2015;43:e47.
- Chen L, Reeve J, Zhang L *et al.* GMPR: a robust normalization method for zero-inflated count data with application to microbiome sequencing data. *PeerJ* 2018;6:e4600.
- Lê Cao K-A, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* 2011;12:253.
- Zhao Y, Cheng M, Zou L *et al.* Hidden link in gut-joint axis: gut microbes promote rheumatoid arthritis at early stage by enhancing ascorbate degradation. *Gut* 2022;71:1041–3.
- Esberg A, Johansson L, Johansson I, Dahlqvist SR. Oral microbiota identifies patients in early onset rheumatoid arthritis. *Microorganisms* 2021;9:1657.
- Tett A, Pasolli E, Masetti G, Ercolini D, Segata N. *Prevotella* diversity, niches and interactions with the human host. *Nat Rev Microbiol* 2021;19:585–99.
- Wells PM, Williams FMK, Matey-Hernandez ML, Menni C, Steves CJ. 'RA and the microbiome: do host genetic factors provide the link?' *J Autoimmun* 2019;99:104–15.
- Alpizar-Rodriguez D, Lesker TR, Gronow A *et al.* *Prevotella copri* in individuals at risk for rheumatoid arthritis. *Ann Rheum Dis* 2019;78:590–3.
- Tett A, Huang KD, Asnicar F *et al.* The *Prevotella copri* complex comprises four distinct clades underrepresented in westernized populations. *Cell Host Microbe* 2019;26:666–79.
- De Filippis F, Pasolli E, Tett A *et al.* Distinct genetic and functional traits of human intestinal *Prevotella copri* strains are associated with different habitual diets. *Cell Host Microbe* 2019;25:444–53.

34. Nii T, Maeda Y, Motooka D *et al*. Genomic repertoires linked with pathogenic potency of arthritogenic *Prevotella copri* isolated from the gut of patients with rheumatoid arthritis. *Ann Rheum Dis* 2023;82:621–9.
35. Fuggle NR, Smith TO, Kaul A, Sofat N. Hand to mouth: a systematic review and meta-analysis of the association between rheumatoid arthritis and periodontitis. *Front Immunol* 2016;7:80.
36. Kitamoto S, Nagao-Kitamoto H, Hein R, Schmidt TM, Kamada N. The bacterial connection between the oral cavity and the gut diseases. *J Dent Res* 2020;99:1021–9.
37. Cook GS, Costerton JW, Lamont RJ. Biofilm formation by *Porphyromonas gingivalis* and *Streptococcus gordonii*. *J Periodontal Res* 1998;33:323–7.
38. Berthelot J-M, Lioté F, Sibilia J. Methotrexate also improves rheumatoid arthritis through correction of microbiota dysbiosis. *Joint Bone Spine* 2023;90:105602.
39. de Arruda AA, Correa JD, Singh Y *et al*. Methotrexate promotes recovery of arthritis-induced alveolar bone loss and modifies the composition of the oral-gut microbiota. *Anaerobe* 2022;75:102577.