

1 **Shotgun metagenomics reveals an enrichment of potentially cross-reactive bacterial epitopes in**
2 **ankylosing spondylitis patients, as well as the effects of TNFi therapy and the host's genotype upon**
3 **microbiome composition.**

4
5 Jian Yin^{1*}, Peter R. Sternes^{2*}, Mingbang Wang³, Mark Morrison⁴, Jing Song¹, Ting Li¹, Ling Zhou¹, Xin Wu¹,
6 Fusheng He⁵, Jian Zhu⁶, Matthew A. Brown^{2,9}, Huji Xu^{1,6,7⁹}.

- 7
8 1. Department of Rheumatology and Immunology, Shanghai Changzheng Hospital, Second Military Medical University,
9 Shanghai, China.
10 2. Translational Genomics Group, Institute of Health and Biomedical Innovation, Queensland University of Technology,
11 Translational Research Institute, Brisbane, Australia.
12 3. Shanghai Key Laboratory of Birth Defects, Division of Neonatology, Children's Hospital of Fudan University, National Center
13 for Children's Health, Shanghai, 201102, China.
14 4. University of Queensland Diamantina Institute, Faculty of Medicine, Translational Research Institute, Brisbane, Australia.
15 5. Immunobio, Shenzhen, Guangdong, 518001, China.
16 6. Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing 100084, China.
17 7. Peking-Tsinghua Center for Life Sciences, Tsinghua University, Beijing 100084, China.

18
19 *These authors contributed equally to this manuscript.
20 ⁹ These authors contributed equally to this manuscript.

21
22 Corresponding Author:
23 Professor Matthew A. Brown,
24 Institute of Health and Biomedical Innovation,
25 Queensland University of Technology,
26 Translational Research Institute,
27 Princess Alexandra Hospital,
28 Woolloongabba, Brisbane,
29 Australia.
30 Telephone: +61 7 3443 7366
31 Fax : +61 7 3443 7102
32 matt.brown@qut.edu.au

33
34 Corresponding Author:
35 Professor Huji Xu
36 Beijing Tsinghua Changgeng Hospital
37 School of Clinical Medicine
38 Tsinghua University, Beijing 100084, China.
39 Phone: +86-10-6279 2998
40 Fax: +86-10-6278 9513
41 email: huji_xu@tsinghua.edu.cn

42 **ABSTRACT**

43 Diverse evidence including clinical, genetic and microbiome studies support a major role of the gut
44 microbiome in the common immune-mediated arthropathy, ankylosing spondylitis (AS). To further
45 investigate this we performed metagenomic analysis of a case-control cohort of 250 Han-Chinese
46 subjects. Previous reports of gut dysbiosis in AS were re-confirmed and several notable bacterial species
47 and functional categories were differentially abundant. TNF-inhibitor (TNFi) therapy at least partially
48 restored the perturbed microbiome observed in untreated AS cases to that of healthy controls, including
49 several important bacterial species that have been previously associated with AS and other related
50 diseases. Enrichment of bacterial peptides homologous to HLA-B27-presented epitopes was observed in
51 the stools of AS patients, suggesting that either HLA-B27 fails to clear these or that they are involved in
52 driving HLA-B27-associated immune reactions. TNFi therapy of AS patients was also associated with a
53 reduction of potentially arthritogenic bacterial peptides, relative to untreated patients. An AS-associated
54 SNP in *RUNX3* significantly influenced the microbiome in two independent cohorts, highlighting a host
55 genotype (other than *HLA-B27*) potentially influencing AS via the microbiome. These findings emphasise
56 the key role that the gut microbiome plays in driving the pathogenesis of AS.

57 INTRODUCTION

58 Ankylosing spondylitis (AS) is a chronic inflammatory disease affecting primarily the spine and pelvis,
59 causing pain and initially reversible stiffness, and ultimately leading to joint ankylosis due to ectopic
60 bone formation. In a subset of patients, peripheral joints and extra articular tissues including the eye,
61 gut and skin are also involved. Its prevalence in Asian and European descent populations ranges from
62 0.09% to 0.55% ^{1,2}, whereas the disease is rare in most of Africa, likely due to the low frequency of the
63 main susceptibility gene, *HLA-B27* ³. There is a significant unmet therapeutic need in AS, with limited
64 evidence that current therapies prevent spinal ankylosis, no oral treatments which suppress disease
65 activity other than corticosteroids, and no treatments which have been demonstrated to induce
66 remission or prevent the disease.

67

68 AS has been shown in both twin and unrelated case/control studies to be highly heritable (twins >90%
69 heritability ^{4,5}, unrelated case/control common variant heritability 69%
70 (<http://www.nealelab.is/blog/2017/9/15/heritability-of-2000-traits-and-disorders-in-the-uk-biobank>)).

71 Over past decade, at least 116 susceptibility genes have been identified, contributing 29% of the overall
72 risk of the disease ⁶. There is substantial evidence suggesting that the interaction between host genetics
73 and gut microbiome is a key driver of the pathogenesis of AS. The high disease heritability indicates that
74 the environmental factors involved in the disease are likely to be ubiquitous. Reactive arthritis is a
75 spondyloarthritis sharing many clinical and genetic features with AS, and is known to be caused by
76 bacterial infections of the gut or urinary tract; a subset of reactive arthritis patients go on to develop AS.
77 About 60% AS patients suffered from subclinical bowel inflammation and 10% of them can be diagnosed
78 as inflammatory bowel disease (IBD) ⁷. There is considerable overlap in the overall heritability of AS and
79 IBD ⁸, the two diseases are often co-familial ⁹, and many shared genetic associations have been
80 identified ¹⁰. A bioinformatic study showed that AS susceptibility genes specifically enriched in gut cells

81 are also enriched in 'response to bacterium' GO term pathway, and that AS-associated genetic loci are
82 found disproportionately to lie within epigenetic marks of gene activity in gut tissue and cells ¹¹.
83 Germ-free *HLA-B27*-transgenic rats and SKG mice are disease-free ^{12,13}. Studies using sequencing-
84 based bacterial profiling of terminal ileal biopsies showed that AS patients have a distinct microbiome ¹⁴,
85 a finding that has subsequently been reproduced studying stool samples in AS patients and patients with
86 spondyloarthritis (SpA), a broader clinical classification ^{15,16}. There has also been suggestive evidence
87 reported that the gut microbiome is associated with differences in AS disease activity ¹⁷. In addition, one
88 study compared SpA patients' stool samples before and three months after TNF-inhibitor (TNFi)
89 treatment onset ¹⁸. Although modest changes were found in microbiome alpha-diversity measures after
90 TNFi treatment, no changes in specific bacterial taxa were observed. This may have been related to
91 power or sampling issues, noting that 15/18 patients studied met only the ASAS axial spondyloarthritis
92 classification criteria rather than having the more specific diagnosis, AS. In summation, the above
93 evidence supports the contention that AS status is influenced by interactions between the gut
94 microbiome and the host immune system.

95

96 To date, the mechanisms involved in the interaction between the host immune system and intestinal
97 microbes remain unclear. One hypothesis suggests that HLA-B27 presents specific peptides to CD8+ T
98 cells, leading to pathogenic adaptive immune responses (the 'arthritogenic peptide theory'). The gut
99 microbiota produce a huge variety and number of peptides, and as such, microbial peptides intrinsic to
100 dysbiosis may activate CD8+ T-cells. In that context, Purcell and colleagues identified 7500 such peptides
101 that bind the eight most common *HLA-B27* subtypes ¹⁹. Here, we present our findings from a shotgun
102 metagenomics sequencing study undertaken with stool samples collected from in 250 Chinese
103 individuals, to investigate evidence of dysbiosis in AS, the effect of host genetic makeup and of TNFi

- 104 treatment on the gut microbiota, and to investigate evidence of immunity to HLA-B27 restricted
105 microbial peptides in AS cases.

106 **MATERIALS AND METHODS**

107 **Subject recruitment**

108 A total of 127 unrelated Han Chinese AS cases were recruited from the Department of Rheumatology
109 and Immunology of Shanghai Changzheng Hospital (Shanghai, China) from December 2014 to June 2017.
110 All cases met the 1984 modified New York criteria for AS ²⁰. 123 healthy controls (blood donors on no
111 prescription medications) were recruited from Shanghai. All human studies have been approved by the
112 Research Ethical Committee of Second Military Medical University, and all patients and controls gave
113 informed written consent for their participation in the studies. Clinical information was recorded for all
114 patients, including demographic information (gender, age, smoking status and BMI), disease duration,
115 *HLA-B27* carriage, sulfasalazine and TNFi treatment information, the Bath Ankylosing Spondylitis Disease
116 Activity Index (BASDAI) ²¹ and Bath Ankylosing Spondylitis Functional Index (BASFI) ²², and clinical
117 manifestations (inflammatory back pain, uveitis, axial arthritis, peripheral arthritis, ulcerative colitis,
118 Crohns disease, enthesitis, dactylitis and psoriasis). Dietary habits were assessed by a 52-question
119 questionnaire to exclude subjects with special dietary habits such as an entirely plant-based or meat-
120 based diet. Where possible, Student's T test and Fisher's exact test were used to identify differences in
121 the metadata categories between cases and controls.

122

123 **DNA microarray and subject genotyping**

124 Samples were genotyped using the Infinium CoreExome-24v1-1 Chip (Illumina, San Diego, CA, USA)
125 according to the manufacturer's recommendations. Bead intensity data were processed and normalised
126 for each sample, and genotypes called within collection using GenomeStudio.

127

128 SNPs with call rate below 95% or with a Hardy-Weinberg equilibrium of $P < 10^{-6}$ in controls were
129 excluded. For the overlapping SNPs, pairwise missingness tests removed all SNPs with differential

130 missingness ($P < 10^{-7}$). After merging data sets, SNPs with call rate below 98% and samples with call rate
131 below 98% were removed. *HLA* alleles were imputed by SNP2HLA using the Pan-Asian reference panel
132 ^{23,24} and SNPs were extracted by PLINK v.1.90 ²⁵.

133

134 **Shotgun metagenome sequencing**

135 Faecal samples were collected and stored at -80°C prior to DNA extraction. DNA was extracted using a
136 StoolGen DNA kit (CWBiotech Co., Beijing, China). DNA concentrations were determined using a Qubit
137 dsDNA BR assay kit (Thermo Fisher, Foster City, CA, USA). 200 – 500 bp insert size libraries were
138 constructed using a TruSeq DNA Sample Preparation Kit (Illumina, San Diego, CA, USA) and an
139 automated SPRI-Works system (Beckman Coulter, San Jose, CA, USA),

140

141 Quality Control (QC) of each library was carried out using an Agilent 2100 Bioanalyzer (Agilent
142 Technologies, Santa Clara, CA, USA), Qubit dsDNA BR assay kit (Thermo Fisher, Foster City, CA, USA) and
143 a KAPA qPCR MasterMix plus Primer Premix kit (Kapa Biosystems, Woburn, MA, USA) according to the
144 manufacturer's instructions. Libraries that passed QC ($>3 \text{ ng}/\mu\text{L}$) were sequenced using an Illumina
145 HiSeq sequencer (Illumina, San Diego, CA, USA) with the paired-end 150-bp sequencing model based on
146 $>5\text{G}$ raw data output per sample.

147

148 Manual inspection and QC of sequencing reads was conducted using FastQC v10.1 ²⁶. Paired-end reads
149 were joined using PEAR v0.9.10 ²⁷ and adapters were trimmed using Trimmomatic v0.36 ²⁸. Contaminant
150 sequences, such as those mapping to human or PhiX genomes, were filtered using Bowtie2 v2.3.4 ²⁹ and
151 the remaining reads were counted and subsampled to an equal sequencing depth of 3,520,000
152 sequencing reads per sample using SeqTK v1.0 ³⁰. MetaPhlan2 v2.6.0 ³¹ was used for taxonomic
153 classification, PanPhlan v1.2.2 ³² was used for strain-level profiling utilising pre-computed pan-genome

154 references where possible, and HUMAnN2 v0.11.1³³ was used for functional mapping to KEGG
155 Orthogroups (KO) and MetaCyc pathways and utilising a UniRef90 database.

156
157 For prediction of bacterial peptides homologous to previously reported HLA-B27-presented epitopes,
158 bacterial-derived sequencing reads were BLASTXed against a local, BLAST-formatted³⁴, version of the
159 immune epitope database (IEDB) v3.0 (downloaded August 2016)^{35,36}. BLAST best-hits with an E-value <
160 0.1 were retained, annotated according to a published study³⁷ and then counted. The peptides
161 annotated as HLA-B27-presented were compared between AS patients and healthy controls using
162 Fisher's exact test.

163

164 **Statistical Analysis**

165 Abundance tables were arcsine square root transformed prior to analysis. Multidimensional data
166 visualisation was conducted using a sparse partial least squares discriminant analysis (sPLSDA) as
167 implemented in R as part of the MixOmics package v6.3.1³⁸, at the species level using Bray-Curtis
168 distance matrices. Receiver operating characteristic curve was calculated from sPLSDA using the
169 MixOmics package v6.3.1. Controlling for covariates (such as gender, BMI, age and smoking status)
170 where appropriate, multivariate association of the bacterial species composition with metadata of
171 interest was conducted using a PERMANOVA test as part of vegan v2.4-5³⁹. The alpha diversity of
172 bacterial species was calculated using the rarefy function, as implemented in vegan v2.4-5. Univariate
173 association of bacterial species and functional pathways/groups were tested for significance using
174 MaAsLin v0.0.5⁴⁰ and Wilcoxon rank-sum tests as implemented in R⁴¹. Only results which were
175 significant in both tests were reported in the main text. For measurement of microbial epitope richness,
176 the Shannon, Simpson and Chao diversity indices were measured (vegan v2.4-5) and group differences
177 were evaluated using Wilcoxon rank-sum tests. Genetic-relatedness dendrograms for strain-level results

178 from PanPhlAn were calculated using Jaccard distance matrices and hierarchical clustering as
179 implemented in R v.3.5.2.

180

181 **RESULTS**

182 **Gut dysbiosis in ankylosing spondylitis**

183 We initially sought to confirm previous reports of dysbiosis in AS cases. The case and control cohorts
184 were divided into discovery and validation cohorts prior to analysis; the discovery cohort consisted of 97
185 AS cases and 93 healthy controls with age-matched demographics, and the remaining 60 subjects
186 comprised the validation cohort (30 AS cases and 30 healthy controls) (Supplementary Table 1). With
187 the exception of a difference in the mean age in the validation cohorts in which the controls were
188 younger on average than cases, no differences were observed between cases and controls in either the
189 discovery or validation cohorts. PERMANOVA and sPLSDA multivariate analysis revealed significant
190 differentiation between the microbial composition of AS cases and healthy controls for both the
191 discovery ($P = 0.019$) and validation ($P = 0.0006$) cohorts (Figure 1A), consistent with previous reports.
192 Receiver-operator curve analysis showed high discrimination between cases and controls using
193 microbiome data alone (AUC=0.87 in combined discovery and validation cohorts) (Supplementary Figure
194 1).

195

196 Seven bacterial species were identified to be differentially abundant ($P < 0.05$) (i.e. were 'indicator
197 species') between AS cases and healthy controls, in both the discovery and validation cohorts (Figure
198 1B). *Clostridiales bacterium 1 7 47FAA*, *Clostridium bolteae* and *Clostridium hatheway* were found to be
199 enriched in AS cases, whilst *Bifidobacterium adolescentis*, *Coprococcus comes*, *Lachnospiraceae*
200 *bacterium 5 1 63FAA* and *Roseburia inulinivorans* were depleted. Several other differentially abundant
201 species of interest were identified in either the discovery or validation cohort, notably *Prevotella copri*,

202 *Dialister invisus* and *Faecalibacterium prausnitzii*. A full list of differentially abundant taxa in either
203 cohort is available in Supplementary Table 2.

204
205 Six KEGG Orthogroups were also found to be differentially abundant ($P < 0.05$) in both cohorts (Figure
206 1C), however there were no MetaCyc metabolic pathways which were differentially abundant in both
207 cohorts. The commonly-differentiated KEGG Orthogroups were EC 2.6.1.9: histidinol-phosphate
208 transaminase, EC 2.7.4.1: polyphosphate kinase, EC 4.3.3.6: pyridoxal 5'-phosphate synthase, EC
209 1.15.1.1: superoxide proteinase, EC 3.4.21.53: ATP-dependent serine phosphatase, and EC 2.4.2.17: ATP
210 phosphoribosyltransferase. Full lists of the differentially abundant KEGG Orthogroups and MetaCyc
211 metabolic pathways are available in Supplementary Tables 3 and 4, respectively.

212
213 Linear regression was used to investigate the correlation between the indicator species and the
214 commonly-differentiated KEGG Orthogroups. All indicator species, except for *Lachnospiraceae*
215 *bacterium 5 1 63FAA*, were significantly associated ($P < 0.05$) with the KEGG Orthogroups, however the
216 degree of variation explained by these species was typically low with R^2 values ranging from 0.0008 to
217 0.13 (0.043 on average) (Supplementary Table 5).

218
219 Strain-level profiling of the dysbiotic microbes identified in Figure 1B uncovered no discernible
220 differences in strain composition between AS cases and healthy controls, with identical strains often
221 being observed in both case and control subjects. (Supplementary Figure 2). This suggests that gut
222 dysbiosis may primarily be a result of differential abundance at the species level and that functional or
223 metabolic differences in the microbiome occur from common genetic elements amongst the strain
224 population, as evidenced by KEGG Orthogroups being detectable in the majority of samples in Figure 1C.

225

226 **Effect of TNFi therapy upon the microbiome**

227 TNFi treatment is highly effective in AS, and it is feasible that at least some of its benefits occur through
228 effects on the gut microbiome. To test this hypothesis, the discovery and validation cohorts were
229 combined into the following categories: healthy controls (n = 123), AS cases treated with TNFi (either
230 etanercept or infliximab, n = 67), and AS cases who have not received TNFi treatment (n = 60). No
231 statistically significant effect of sulfasalazine treatment was observed (P=0.76, Supplementary Figure 3).
232 Multivariate comparison of TNFi untreated and treated cases revealed an effect of TNFi treatment upon
233 the overall composition of the microbiome (P = 0.022) (Figure 2A). Untreated cases were significantly
234 different to healthy controls (P = 0.0002), whereas treated cases were not significantly different to
235 healthy controls (P = 0.069) indicating that treatment has helped restore the perturbed composition of
236 the microbiome.

237
238 To identify the key species modulated by the effects of TNFi therapy, species which were both (a)
239 perturbed in untreated AS cases relative to healthy controls, and (b) differently abundant in treated
240 cases compared to untreated cases, were first identified (Figure 2B and Supplementary Table 6). Six of
241 the eight identified species exhibited significant depletion in untreated AS cases, however TNFi
242 treatment appeared to restore the abundance of these species to levels indistinguishable from healthy
243 controls. These species were: *Prevotella copri*, *Faecalibacterium prausnitzii*, *Bilophila unclassified*,
244 *Klebsiella pneumoniae*, *Ruminococcus bromii* and *Eubacterium bifforme*. The remaining two species
245 (*Clostridium symbiosum* and *Eggerthella unclassified*) were enriched in untreated AS and their
246 abundance was no longer different to healthy controls in treated cases. The findings in relation to
247 *Prevotella copri* and *Klebsiella pneumoniae* were of particular interest given their previous association
248 with rheumatoid arthritis (RA) and AS, respectively, as was the highly abundant (approximately 20% of
249 total bacterial DNA, on average) *Faecalibacterium prausnitzii* for its notable depletion in several

250 autoimmune diseases⁴². TNFi therapy appeared to partially normalise the dysbiotic bacterial species
251 and KEGG Orthogroups observed in AS cases relative to healthy controls shown in Figures 1B and 1C,
252 however no statistically significant differences between treated and untreated cases were observed,
253 potentially due to sample size constraints (Supplementary Figure 4).

254
255 The above approach was also used to identify metabolic pathways modulated by TNFi therapy. 20
256 MetaCyc metabolic pathways were identified in total and the perturbed abundance observed in
257 untreated AS cases was restored to healthy control levels in 17 of these. In broad terms, these pathways
258 primarily related to amino acid biosynthesis (notably branched-chain and aromatic amino acid
259 biosynthesis), carbohydrate metabolism (notably starch degradation), nucleotide biosynthesis,
260 metabolite biosynthesis and cell structure. Specific details of the 20 differentially abundant MetaCyc
261 pathways are available in Supplementary Table 7.

262
263 Linear regression was used to investigate the association of between the modulated species and
264 modulated pathways (Supplementary Table 8). Except for PWY-6545: Pyrimidine biosynthesis which was
265 not associated with any individual identified species, all the pathways were significantly associated with
266 at least two of the identified species. Similarly, all the species were significantly associated with multiple
267 pathways, however the abundances of *Bilophila unclassified* and *Klebsiella pneumoniae* were inversely
268 correlated with pathway abundance. An increase in *Klebsiella pneumoniae* was associated with a
269 decrease in the abundance of PWY-6737: starch degradation ($P = 0.014$; $R^2 = -0.0426$). The observed
270 decrease in the starch degradation pathway for untreated AS cases is primarily attributed to a depletion
271 of *Faecalibacterium prausnitzii* ($P = 2.38 \times 10^{-24}$; $R^2 = 0.3123$). *Faecalibacterium prausnitzii* also exhibited
272 strong associations with other metabolic pathways.

273

274 Strain-level profiling of the bacterial species outlined in Figure 2B also revealed no discernible
275 differences in strain composition between healthy controls, treated cases and untreated cases,
276 indicating the TNFi therapy affected the relative abundance of each species, not necessarily the
277 underlying strain composition (Supplementary Figure 5).

278

279 **Effect of host genotype upon the microbiome**

280 Genome-wide association studies (GWAS) have identified many genetic loci which are associated with
281 AS. Emerging evidence indicates that alleles such as *HLA-B27* may influence the disease through its
282 effect upon the gut microbiome^{43,44}. To investigate whether additional loci may affect the gut
283 microbiome and potentially influence disease, we performed PERMANOVA analysis upon loci known to
284 be associated with AS.

285

286 Considering non-MHC loci, an association was noted for rs11249215, a SNP in the promoter of runt-
287 related transcription factor 3 (*RUNX3*) gene known to be associated with AS^{45,46}. This variant was
288 associated with a shift in the microbiome of both AS cases and healthy controls (combined $P = 0.0097$).
289 Furthermore, sPLSDA revealed that the degree alteration appears dependent on whether the host
290 carried a heterozygous or homozygous genotype (Figure 3A), with the homozygous genotype resulting in
291 a more substantial shift. As further confirmation, we analysed a recently published 16S metabarcoding
292 dataset⁴⁷ of 107 healthy control subjects which were sampled from six different body sites. This analysis
293 re-confirmed discrimination of the microbiomes based on genotype (PERMANOVA; $P = 0.0001$) (Figure
294 3B). The *RUNX3* SNP had no observable effect upon the dysbiotic bacterial species and KEGG
295 Orthogroups outlined in Figures 1B and 1C, however its effects upon species richness and community
296 composition (Figure 3A) suggest that further research is required to confirm a role in AS pathogenesis
297 via effects upon the microbiome. The effect of *HLA-B27* upon the microbiome of the current cohort was

298 unable to be investigated due to high prevalence of this genotype in AS cases and low prevalence in
299 healthy controls, thus the effect of *HLA-B27* was unable to be discerned from the effect of AS itself.

300

301 To highlight the key bacterial species affected by rs11249215, species associated with the heterozygous
302 (AG) and homozygous (GG) genotypes, or solely the homozygous (GG) genotype were identified (Figure
303 3C). Several bacterial species showed differential abundance for the AG genotype, but not the GG
304 genotype, potentially due to sample size constraints (the GG genotype was present in 37 of the 188
305 genotyped subjects), and thus were excluded from further analysis. Four key species identified as
306 depleted in the AG/GG genotypes were: *Lachnospiraceae bacterium 1 1 57FAA*, *Eubacterium*
307 *ventriosum*, *Citrobacter freundii* and *Citrobacter unclassified* (Supplementary Table 9).

308

309 Six MetaCyc metabolic pathways were found to be differentially abundant when comparing *RUNX3*
310 rs11249215 genotypes (Figure 3D). Similar to the differences found for TNFi therapy, the differential
311 pathways were primarily associated with amino acid and nucleotide biosynthesis, however notable
312 differences in polyamine biosynthesis and pyruvate fermentation (to acetate and lactate) were also
313 observed. Of interest is the polyamine biosynthesis pathway for the role of polyamines in enhancing the
314 integrity of the intestinal epithelial cell barrier, and the adenosine biosynthesis pathway for the anti-
315 inflammatory and immunosuppressant effect of adenosine (Supplementary Table 10)^{48,49}.

316

317 Linear regression of these species with the metabolic pathways revealed relatively marginal
318 associations, with the two *Citrobacter* species exhibiting no association with any metabolic pathway
319 (Supplementary Table 11). 16S rRNA gene metabarcoding analysis of the predominately Caucasian
320 cohort sampled from various body sites (Figure 3B) revealed a different set of taxa and metabolic
321 pathways potentially influenced by the *RUNX3* SNP (Supplementary Tables 12 and 13). The minimal

322 overlap with the current shotgun metagenomic study is potentially indicative of the differences between
323 the metagenomic approaches and/or differences in studied cohorts (i.e. geographic location, diet,
324 ethnicity...etc).

325

326 Similar to the strain-level results for AS status and TNFi therapy, no observable bias in the underlying
327 strain population was observed, indicating that *RUNX3* variants likely affect the relative abundance of
328 species, not necessarily strain composition (Supplementary Figure 6^{47,50}). Comparatively fewer species
329 were associated with the *RUNX3* SNP in comparison to the number of species associated with AS status
330 and TNFi treatment. Consistent with recent publications which have investigated the effect of the host's
331 genotype upon the abundance of specific taxa^{47,50}, these data provide supporting evidence that the
332 underlying host's genetics may have a generalised effect upon the microbiome, with a subtle effect on a
333 higher number of taxa as opposed to a marked effect on a select few.

334

335 **Bacterial-derived HLA-B27 epitopes in AS cases and healthy controls**

336 The main physiological function of HLA-B27 is to present peptides to CD8 lymphocytes, thereby driving
337 cell mediated immune reactions. Differences in the presence of HLA-B27-positive epitopes in the gut
338 microbiome in cases compared to controls, and in HLA-B27 carriers compared with HLA-B27-negative
339 subjects, would be consistent with effects of HLA-B27 to 'shape' the gut microbiome, and the
340 significance of this in regards disease pathogenesis.

341

342 To investigate the abundance of bacterial peptides homologous to HLA-B27 epitopes in AS cases and
343 healthy controls, translated nucleotide searches were performed against IEBD v3.0, annotated according
344 to a published study³⁷ and counted. Significant enrichment of these peptide sequences was observed in
345 AS cases, with 24 of these enriched in both the discovery and validation cohorts (Table 1). AS cases not

346 only exhibited enrichment of these peptides but the overall diversity of peptides was increased, with
347 Shannon, Inverse Simpson and Chao diversity indices revealing significant differences between AS cases
348 and healthy controls (Figure 4A).

349

350 TNFi treatment effects were also investigated. The overall abundance and diversity of bacterial peptides
351 homologous to HLA-B27-presented epitopes was significantly different between the different treatment
352 categories (Figure 4B). Untreated AS cases exhibited increased abundance and diversity of peptides. For
353 patients who underwent TNFi therapy, a reduction in these potentially arthritogenic peptides was
354 observed relative to untreated cases, however their levels remained marginally higher than healthy
355 controls.

356 **DISCUSSION**

357 **Gut dysbiosis in ankylosing spondylitis**

358 This study re-confirmed the occurrence of bacterial gut dysbiosis in AS cases and identified seven
359 bacterial species which were commonly differentiated between cases and controls in both the discovery
360 and validation cohorts (Figure 1B). Two of these species, *Bifidobacterium adolescentis* and *Coprococcus*
361 *comes*, have been noted for their depletion in Crohn's Disease⁵¹ and were also observed to be depleted
362 in AS cases in this study. An additional two species previously reported to be associated with AS,
363 *Prevotella copri* and *Dialister invisus*, were found to be differentially abundant in either the discovery or
364 validation cohorts (Supplementary Table 2). In the case of *Prevotella copri*, previous studies have
365 demonstrated enrichment in new onset RA cases yet depletion in chronic RA cases⁵². Consistent with
366 these findings, our study found that *Prevotella copri* was depleted in AS cases within the non-age-
367 matched cohort, for which the demographics were heavily skewed towards older AS patients with long-
368 standing disease (Supplementary Table 1). Previous studies in AS have shown increases in *Prevotellaceae*
369¹⁴, or specifically with this species¹⁵. As discussed below, *Prevotella copri* carriage normalised with TNFi
370 treatment. Further studies will be required to determine if *Prevotella copri* carriage changes with
371 disease duration, as has been reported in RA.

372

373 Carriage of *Dialister* species has been previously associated with disease activity in spondylarthrosis
374 patients¹⁷, but the carriage of *Dialister invisus* has been reported to be decreased in inflammatory
375 bowel disease (IBD)^{53,54}. Whilst we found enrichment of *Dialister invisus* in AS cases in the discovery
376 cohort, this was not confirmed in the validation cohort, nor has it been reported in other AS studies. Its
377 pathogenic significance is therefore uncertain.

378

379 Of particular interest, the notable ‘peace keeping’ microbe *Faecalibacterium prausnitzii* was also found
380 to be depleted in AS cases in the validation cohort. This bacterium has also been consistently shown to
381 be depleted in IBD⁵⁴⁻⁵⁹, and has been previously shown to be depleted in the disease enthesitis-related
382 arthritis, a paediatric disease-classification which includes children with ankylosing spondylitis⁶⁰.
383 *Faecalibacterium prausnitzii* is known to produce butyrate and other metabolites and peptides that have
384 diverse anti-inflammatory effects including promoting T-regulatory cell differentiation⁶¹, influences on
385 Th17 lymphocyte activation, and promotion of gut mucosal barrier function^{59,62,63}. As discussed below,
386 TNFi treatment also led to normalisation of *Faecalibacterium prausnitzii* carriage. These findings suggest
387 that *Faecalibacterium prausnitzii* plays a key anti-inflammatory role in AS, as it does in IBD.

388
389 The power of metagenome sequencing is to augment the widely observed phenomena collectively
390 referred to as “dysbiosis” beyond taxonomy-based assessment of microbiome, and provide a more
391 highly resolved and functional characterisation of the microbiome. Here, variation in several KEGG
392 Orthogroups remained consistent between the discovery and validation cohorts (Figure 1C).
393 Additionally, differential abundance of some MetaCyc metabolic pathways was also observed
394 (Supplementary Table 4), but these differences were not consistent between discovery and validation
395 cohorts; potentially highlighting the confounding influence of the host’s age upon the metabolic
396 composition of the gut microbiome.

397
398 Interestingly, genes encoding pyridoxal 5'-phosphate synthase, an important enzyme for the
399 biosynthesis of vitamin B6, were less abundant amongst the microbiome of AS cases compared to
400 healthy controls in both cohorts. Vitamin B6 plays a role in the maintenance of vitamin homeostasis in
401 colonocytes^{64,65}. It has been found to modulate colonic inflammation and several studies have
402 investigated the role of vitamin B6 for the treatment of inflammation in RA patients⁶⁶⁻⁷¹. Evidence from

403 case-control studies show that RA patients have low vitamin B6 status compared to healthy controls,
404 however intervention studies have yielded inconsistent findings, possibly due to the dose of the
405 administered vitamin B6. The reduced potential for the microbiome of AS patients to produce pyridoxal
406 5'-phosphate synthase, and thus vitamin B6, may warrant further investigation into intervention
407 strategies to mitigate inflammation in AS patients.

408

409 **Effect of TNFi therapy upon the microbiome**

410 Previous study of RA patients before and after synthetic disease-modifying anti-rheumatic drug
411 treatment revealed moderate differences in the gut microbiota composition, with the perturbed
412 microbial composition being partly restored following treatment⁷². Similarly, analysis of spondylarthrosis
413 patients before and after TNFi therapy also revealed modest differences in microbial composition yet no
414 specific taxon was found to be modulated, likely due to sample size⁷³. Utilising a larger sample size, we
415 confirmed that TNFi therapy was correlated with a restoration of the perturbed microbial composition,
416 and additionally identified several notable bacterial species modulated by treatment.

417

418 We observed that TNFi therapy restored the depletion of *Faecalibacterium prausnitzii* in AS cases.
419 Restoration of *Faecalibacterium prausnitzii* abundance was also correlated with the restored abundance
420 of aromatic and branched-chain amino acid biosynthesis pathways. A recent ulcerative colitis study⁵⁵
421 revealed reduced dysbiosis and increased *Faecalibacterium prausnitzii* abundance in responders
422 compared with non-responders following TNFi therapy. Furthermore, recovery of *Faecalibacterium*
423 *prausnitzii* in patients with ulcerative colitis after relapse was associated with maintenance of remission
424⁷⁴. Another study demonstrated that treatment of infliximab completely restored *Faecalibacterium*
425 *prausnitzii* concentrations from zero to 1.4×10^{10} bacteria/mL within few days⁵².

426

427 Another important microbe, *Prevotella copri*, was observed to be enriched to levels closely matched to
428 that of healthy controls following TNFi treatment. Abundance of *Prevotella copri* has previously been
429 shown to be enriched in untreated new onset RA patients yet depleted in chronic RA cases, patients
430 with psoriatic arthritis and healthy controls⁵². Colonisation of SKG mice with *Prevotella copri*-dominated
431 microbiota from RA patients exhibited increased number of Th17 cells in the large intestine⁷⁵. HLA-DR-
432 presented peptides (T cell epitopes) from *Prevotella copri* were recently found to stimulate Th1
433 responses in 42% of new onset RA cases, with subgroups of RA patients demonstrating differential IgG
434 or IgA immune reactivity, providing evidence that *Prevotella copri* is immune-relevant in RA
435 pathogenesis⁷⁶. Additionally, the presence of the *HLA-DRB1* risk allele, which influences disease
436 severity, in RA patients was found to be inversely correlated with *Prevotella copri* abundance^{52,77-79}. A
437 recent study of Chinese AS patients revealed enrichment of *Prevotella copri*, as well as *Prevotella*
438 *melaninogenica* and *Prevotella* sp. C561^{15,80}. Contrasting these results, in the current study we observed
439 depletion of *Prevotella copri* in untreated AS cases, which was restored to the healthy control levels in
440 TNFi-treated patients. These seemingly conflicting reports of *Prevotella copri* abundance may be
441 explained by the large degree of intraspecific genetic diversity of *Prevotella copri* strains, with strain
442 variation adding an additional layer of complexity for predicting the function of *Prevotella copri* in the
443 gut. The *Prevotella* genus also contains members that may be beneficial, and which do not function as
444 pathobionts⁷⁷⁻⁷⁹, with observed enrichment in healthy individuals. Taken together, our results which
445 demonstrate a modulation of *Prevotella copri* abundance in TNFi-treated cases is a noteworthy
446 observation, however without a stronger grasp of the strain-level genome variation within this taxon
447 and their prevalence across our cohort, attempts to therapeutically modulate and predict the effects of
448 *Prevotella copri* remains a significant challenge⁸¹.

449

450 *Klebsiella pneumoniae* has also been suggested to play a significant role in AS pathogenesis⁸², although
451 this remains controversial⁸³. *Klebsiella pneumoniae* notably produces pullulanase, a starch-debranching
452 enzyme which enables the degradation of starch into simple sugars⁸⁴. The apparent arthritogenic
453 effects of dietary starch in AS are based on the concept that the growth of *Klebsiella sp.* are favoured by
454 these diets and drive AS pathogenesis^{85,86}. Consequently, low starch diets have been promoted and are
455 frequently followed by patients⁸⁶, although there is no published evidence to date as to their efficacy in
456 positively affecting AS disease course. Here, we actually observed a depletion of this microbe in
457 untreated cases relative to healthy controls, whereas TNFi-treated cases showed a restoration of this
458 bacterium. Furthermore, our metagenome sequencing data showed an inverse correlation between
459 *Klebsiella pneumoniae* relative abundance and the overall starch degradation metabolic pathway ($P =$
460 0.014 ; $R^2 = -0.043$) (Supplementary Table 5). This pathway not only includes the pullulanase-mediated
461 starch de-branching reaction, but also further downstream reactions including the transport and
462 catabolism of maltodextrins. These findings do not support an association between *Klebsiella*
463 *pneumoniae* and AS pathophysiology, although the role of dietary and/or resistant starches on the gut
464 microbiota and AS warrants further investigation.

465

466 MetaCyc pathway analysis revealed depletion of the aromatic and branched-chain amino acid
467 biosynthesis pathways in untreated AS cases, which are responsible for production of four of the nine
468 essential amino acids in humans (leucine, isoleucine, valine and phenylalanine) (Supplementary Table 7).
469 Vitamin B6 is an essential co-factor for branched-chain amino acid transaminase, the last step of
470 branched-chain amino acid synthesis, and was found to be depleted in RA cases, as previously
471 mentioned. Therefore, in the current study we not only observed a depletion of genes encoding the
472 branched-chain amino acid biosynthesis pathway, but also for the enzyme which synthesises an
473 important co-factor in the process.

474

475 **Enrichment of potentially arthritogenic bacterial peptides**

476 Pathogenic bacteria have long been hypothesised as an immunological triggers of AS pathogenesis. In
477 the current study, patients with AS not only demonstrated an enrichment of bacterial peptides matching
478 HLA-B27 epitopes (Table 1), but the diversity of these peptides was greater overall (Figure 4A). These
479 data provide supporting evidence for the molecular mimicry hypothesis for which bacterial-derived
480 peptides may stimulate AS via cross-activation of autoreactive T- or B- cells, thus leading to
481 autoimmunity. This hypothesis does not however explain the increase in HLA-B27 epitopes amongst AS
482 cases, which could be explained by effects of non-HLA genetic factors, or AS-associated environmental
483 factors. An alternate hypothesis is that their excess carriage is caused by a deficiency in the ability of
484 HLA-B27 to effectively control their presence, consistent with evidence of increased bacterial migration
485 across the gut mucosa in AS ⁸⁷. Interestingly, the modulation of the gut microbiome caused by TNFi
486 treatment restored the elevated abundance and diversity of peptides observed in untreated cases to
487 levels which were more closely matched to healthy controls (Figure 4B). Further research will be
488 required to resolve these alternate hypotheses.

489

490 **Effect of host genotype upon the microbiome**

491 Very recently, it was demonstrated that HLA-B27 is associated with a significant shift in the microbiome
492 in healthy individuals ⁴⁷. Furthermore, in mouse models, MHC polymorphisms were demonstrated to
493 contribute to an individual's microbial composition, thus influencing health ⁸⁸. Our investigation
494 revealed an additional AS-associated SNP, rs11249215 in *RUNX3* ^{45,46}, which was also correlated with a
495 significant shift in bacterial composition (Figure 3A). This result was replicated in a confirmatory dataset
496 of healthy Caucasian individuals (Figure 3B). In addition to potential roles in autoimmune diseases,
497 variants in *RUNX3* have been associated with the intestinal inflammatory disorder celiac disease ⁸⁹, and

498 *RUNX3*-knockout mice spontaneously develop IBD ⁹⁰. It is therefore tempting to hypothesise that the
499 role of *RUNX3* in disease pathogenesis is, at least in part, caused by perturbation of the gut microbiome.
500 Interestingly, subjects homozygous for rs11249215 exhibited a significant decrease in the abundance of
501 the polyamine biosynthesis superpathway (Figure 3D). The intestinal tract contains high levels of
502 polyamines which are critical for cell growth and can stimulate the production of junction proteins
503 which are crucial for regulating paracellular permeability and reinforcing epithelial barrier function.
504 Shifts in host and microbial polyamine metabolism may also alter the cytokine environment and induce
505 cellular processes in both acute and chronic inflammatory settings ⁴⁹. A potential relationship between
506 *RUNX3*, microbial composition, intestinal polyamine levels and epithelial permeability and/or the
507 cytokine environment warrants further investigation.
508

509 **CONCLUSION**

510 In this study we confirm that AS is characterised by gut dysbiosis and identify key indicator species,
511 several of which are shared with IBD. This dysbiosis is associated with functional differences in the
512 microbiome involving known inflammation-related pathways. We demonstrate that treatment with
513 TNFi, which is highly effective in suppressing the clinical manifestations of AS, normalises the gut
514 microbiome, and its functional properties, in AS cases. We further demonstrate that the AS gut
515 microbiome is enriched for bacterial peptides that have previously been shown to be presented by HLA-
516 B27, and that this enrichment is also normalised by TNFi treatment. The impact of the host's genotype
517 upon microbiome composition was also highlighted, with an AS- and IBD-associated SNP in *RUNX3*
518 correlating with a shift in microbiome composition. These findings are consistent with disease models in
519 which AS pathogenesis is driven by interactions between a genetically primed host immune system, and
520 the gut microbiome, and point to potential therapeutic and/or preventative approaches for the disease.

521

522 **AUTHOR CONTRIBUTIONS**

523 Study design was performed by HX, JY and MAB. Subject recruitment and sample collection was
524 performed by JY, JS, TL, LZ, XW and JZ. Metagenomic analysis was performed by PRS, and bacterial
525 epitope studies by JY, FH and MW. The manuscript was prepared by PRS, MM, MAB and HX.

526

527 **ACKNOWLEDGEMENTS**

528 HX is supported by National Science Foundation of China (Grant 81302578) and China Ministry of
529 Science and technology (973 Program of China 2014CB541800). MAB is funded by a National Health and
530 Medical Research Council Senior Principal Research Fellowship (APP1024879).

531

532 REFERENCES

- 533 1. Wang, R. & Ward, M.M. Epidemiology of axial spondyloarthritis: an update. *Current opinion in*
534 *rheumatology* **30**, 137-143 (2018).
- 535 2. Braun, J., *et al.* Prevalence of spondylarthropathies in HLA-B27 positive and negative blood
536 donors. *Arthritis Rheum* **41**, 58-67 (1998).
- 537 3. Brown, M.A., *et al.* Ankylosing spondylitis in West Africans--evidence for a non-HLA-B27
538 protective effect. *Ann Rheum Dis* **56**, 68-70 (1997).
- 539 4. Pedersen, O.B., *et al.* Ankylosing spondylitis in Danish and Norwegian twins: occurrence and the
540 relative importance of genetic vs. environmental effectors in disease causation. *Scand J*
541 *Rheumatol* **37**, 120-126 (2008).
- 542 5. Brown, M.A., *et al.* Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and
543 the environment. *Arthritis Rheum* **40**, 1823-1828 (1997).
- 544 6. Ranganathan, V., Gracey, E., Brown, M.A., Inman, R.D. & Haroon, N. Pathogenesis of ankylosing
545 spondylitis - recent advances and future directions. *Nature reviews. Rheumatology* **13**, 359-367
546 (2017).
- 547 7. Mielants, H., *et al.* The evolution of spondyloarthropathies in relation to gut histology. II.
548 Histological aspects. *J Rheumatol* **22**, 2273-2278 (1995).
- 549 8. Ellinghaus, D., *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations
550 and highlights disease-specific patterns at shared loci. *Nature genetics* **48**, 510-518 (2016).
- 551 9. Thjodleifsson, B., Geirsson, A.J., Bjornsson, S. & Bjarnason, I. A common genetic background for
552 inflammatory bowel disease and ankylosing spondylitis: a genealogic study in Iceland. *Arthritis*
553 *Rheum* **56**, 2633-2639 (2007).
- 554 10. Parkes, M., Cortes, A., van Heel, D.A. & Brown, M.A. Genetic insights into common pathways
555 and complex relationships among immune-mediated diseases. *Nature reviews. Genetics* **14**, 661-
556 673 (2013).
- 557 11. Li, Z., *et al.* Epigenetic and gene expression analysis of ankylosing spondylitis-associated loci
558 implicate immune cells and the gut in the disease pathogenesis. *Genes and immunity* **18**, 135-
559 143 (2017).
- 560 12. Taurog, J.D., *et al.* The germfree state prevents development of gut and joint inflammatory
561 disease in HLA-B27 transgenic rats. *The Journal of experimental medicine* **180**, 2359-2364 (1994).
- 562 13. Rehaume, L.M., *et al.* ZAP-70 genotype disrupts the relationship between microbiota and host,
563 leading to spondyloarthritis and ileitis in SKG mice. *Arthritis & rheumatology* **66**, 2780-2792
564 (2014).
- 565 14. Costello, M.E., *et al.* Brief Report: Intestinal Dysbiosis in Ankylosing Spondylitis. *Arthritis &*
566 *rheumatology* **67**, 686-691 (2015).
- 567 15. Wen, C., *et al.* Quantitative metagenomics reveals unique gut microbiome biomarkers in
568 ankylosing spondylitis. *Genome biology* **18**, 142 (2017).
- 569 16. Breban, M., *et al.* Faecal microbiota study reveals specific dysbiosis in spondyloarthritis. *Ann*
570 *Rheum Dis* **76**, 1614-1622 (2017).
- 571 17. Tito, R.Y., *et al.* Brief Report: Dialister as a Microbial Marker of Disease Activity in
572 Spondyloarthritis. *Arthritis & rheumatology* **69**, 114-121 (2017).
- 573 18. Bazin, T., *et al.* Microbiota Composition May Predict Anti-Tnf Alpha Response in
574 Spondyloarthritis Patients: an Exploratory Study. *Scientific reports* **8**, 5446 (2018).
- 575 19. Schittenhelm, R.B., Sian, T.C., Wilmann, P.G., Dudek, N.L. & Purcell, A.W. Revisiting the
576 arthritogenic peptide theory: quantitative not qualitative changes in the peptide repertoire of
577 HLA-B27 allotypes. *Arthritis & rheumatology* **67**, 702-713 (2015).

- 578 20. Linden, S.V.D., Valkenburg, H.A. & Cats, A. Evaluation of diagnostic criteria for ankylosing
579 spondylitis. *Arthritis & rheumatology* **27**, 361-368 (1984).
- 580 21. Garrett, S., *et al.* A new approach to defining disease status in ankylosing spondylitis: the Bath
581 Ankylosing Spondylitis Disease Activity Index. *J Rheumatol* **21**, 2286-2291 (1994).
- 582 22. Calin, A., *et al.* A new approach to defining functional ability in ankylosing spondylitis: the
583 development of the Bath Ankylosing Spondylitis Functional Index. *J Rheumatol* **21**, 2281-2285
584 (1994).
- 585 23. Das, S., *et al.* Next-generation genotype imputation service and methods. *Nature genetics* **48**,
586 1284 (2016).
- 587 24. Jia, X., *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PloS one* **8**,
588 e64683 (2013).
- 589 25. Rentería, M.E., Cortes, A. & Medland, S.E. Using PLINK for genome-wide association studies
590 (GWAS) and data analysis. in *Genome-Wide Association Studies and Genomic Prediction* 193-213
591 (Springer, 2013).
- 592 26. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).
- 593 27. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End
594 reAd mergeR. *Bioinformatics* **30**, 614-620 (2013).
- 595 28. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
596 data. *Bioinformatics* **30**, 2114-2120 (2014).
- 597 29. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357
598 (2012).
- 599 30. Li, H. seqtk Toolkit for processing sequences in FASTA/Q formats. (GitHub, 2012).
- 600 31. Truong, D.T., *et al.* MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature*
601 *methods* **12**, 902 (2015).
- 602 32. Scholz, M., *et al.* Strain-level microbial epidemiology and population genomics from shotgun
603 metagenomics. *Nature methods* **13**, 435 (2016).
- 604 33. Abubucker, S., *et al.* Metabolic reconstruction for metagenomic data and its application to the
605 human microbiome. *PLoS computational biology* **8**, e1002358 (2012).
- 606 34. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool.
607 *Journal of molecular biology* **215**, 403-410 (1990).
- 608 35. Vita, R., *et al.* The immune epitope database (IEDB) 3.0. *Nucleic acids research* **43**, D405-D412
609 (2014).
- 610 36. Kim, Y., *et al.* Immune epitope database analysis resource. *Nucleic acids research* **40**, W525-
611 W530 (2012).
- 612 37. Schittenhelm, R.B., Sian, T.C., Wilmann, P.G., Dudek, N.L. & Purcell, A.W. Revisiting the
613 arthritogenic peptide theory: quantitative not qualitative changes in the peptide repertoire of
614 HLA-B27 allotypes. *Arthritis & rheumatology* **67**, 702-713 (2015).
- 615 38. Lê Cao, K.-A., *et al.* MixMC: a multivariate statistical framework to gain insight into microbial
616 communities. *PloS one* **11**, e0160169 (2016).
- 617 39. Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*
618 **14**, 927-930 (2003).
- 619 40. Morgan, X.C., *et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and
620 treatment. *Genome biology* **13**, R79 (2012).
- 621 41. Ihaka, R. & Gentleman, R. R: a language for data analysis and graphics. *Journal of computational*
622 *and graphical statistics* **5**, 299-314 (1996).
- 623 42. Forbes, J.D., Van Domselaar, G. & Bernstein, C.N. The gut microbiota in immune-mediated
624 inflammatory diseases. *Frontiers in microbiology* **7**, 1081 (2016).

- 625 43. Kenna, T.J. & Brown, M.A. Immunopathogenesis of ankylosing spondylitis. *Int J Clin Rheumatol* **8**,
626 265-274 (2013).
- 627 44. Rosenbaum, J.T., *et al.* Does the microbiome play a causal role in spondyloarthritis? *Clinical*
628 *rheumatology* **33**, 763-767 (2014).
- 629 45. Evans, D.M., *et al.* Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates
630 peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nature genetics* **43**,
631 761 (2011).
- 632 46. International Genetics of Ankylosing Spondylitis, C., *et al.* Identification of multiple risk variants
633 for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nature*
634 *genetics* **45**, 730-738 (2013).
- 635 47. Asquith, M., *et al.* HLA alleles associated with risk of ankylosing spondylitis and rheumatoid
636 arthritis influence the gut microbiome. *bioRxiv*, 517813 (2019).
- 637 48. Varani, K., *et al.* The role of adenosine receptors in rheumatoid arthritis. *Autoimmunity reviews*
638 **10**, 61-64 (2010).
- 639 49. Rooks, M.G. & Garrett, W.S. Gut microbiota, metabolites and host immunity. *Nature Reviews*
640 *Immunology* **16**, 341 (2016).
- 641 50. Rothschild, D., *et al.* Environment dominates over host genetics in shaping human gut
642 microbiota. *Nature* **555**, 210 (2018).
- 643 51. Gevers, D., *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell host &*
644 *microbe* **15**, 382-392 (2014).
- 645 52. Scher, J.U., *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility
646 to arthritis. *elife* **2**(2013).
- 647 53. Joossens, M., *et al.* Dysbiosis of the faecal microbiota in patients with Crohn's disease and their
648 unaffected relatives. *Gut* **60**, 631-637 (2011).
- 649 54. Schirmer, M., *et al.* Dynamics of metatranscription in the inflammatory bowel disease gut
650 microbiome. *Nat Microbiol* **3**, 337-346 (2018).
- 651 55. Magnusson, M.K., *et al.* Anti-TNF therapy response in patients with ulcerative colitis is
652 associated with colonic antimicrobial peptide expression and microbiota composition. *Journal of*
653 *Crohn's and Colitis* **10**, 943-952 (2016).
- 654 56. De Preter, V., *et al.* Faecal metabolite profiling identifies medium-chain fatty acids as
655 discriminating compounds in IBD. *Gut*, gvtjnl-2013-306423 (2014).
- 656 57. Le Gall, G., *et al.* Metabolomics of fecal extracts detects altered metabolic activity of gut
657 microbiota in ulcerative colitis and irritable bowel syndrome. *Journal of proteome research* **10**,
658 4208-4218 (2011).
- 659 58. Marchesi, J.R., *et al.* Rapid and noninvasive metabonomic characterization of inflammatory
660 bowel disease. *Journal of proteome research* **6**, 546-551 (2007).
- 661 59. Machiels, K., *et al.* A decrease of the butyrate-producing species *Roseburia hominis* and
662 *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut* **63**, 1275-
663 1283 (2014).
- 664 60. Stoll, M.L., *et al.* Altered microbiota associated with abnormal humoral immune responses to
665 commensal organisms in enthesitis-related arthritis. *Arthritis Res Ther* **16**, 486 (2014).
- 666 61. Arpaia, N., *et al.* Metabolites produced by commensal bacteria promote peripheral regulatory T-
667 cell generation. *Nature* **504**, 451-455 (2013).
- 668 62. Sokol, H., *et al.* *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium
669 identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* **105**,
670 16731-16736 (2008).
- 671 63. Neis, E., Dejong, C. & Rensen, S.J.N. The role of microbial amino acid metabolism in host
672 metabolism. *7*, 2930-2946 (2015).

- 673 64. Rucker, R.B., Suttie, J.W. & McCormick, D.B. *Handbook of vitamins*, (CRC Press, 2001).
- 674 65. Magnúsdóttir, S., Ravcheev, D., de Crécy-Lagard, V. & Thiele, I. Systematic genome assessment
675 of B-vitamin biosynthesis suggests co-operation among gut microbes. *Frontiers in genetics* **6**,
676 148 (2015).
- 677 66. Woolf, K. & Manore, M.M. Elevated plasma homocysteine and low vitamin B-6 status in
678 nonsupplementing older women with rheumatoid arthritis. *Journal of the American Dietetic*
679 *Association* **108**, 443-453 (2008).
- 680 67. Schumacher, H., Bernhart, F. & György, P. Vitamin B6 levels in rheumatoid arthritis: effect of
681 treatment. *The American journal of clinical nutrition* **28**, 1200-1203 (1975).
- 682 68. Chiang, E.-P.I., Bagley, P.J., Selhub, J., Nadeau, M. & Roubenoff, R. Abnormal vitamin B6 status is
683 associated with severity of symptoms in patients with rheumatoid arthritis. *The American*
684 *journal of medicine* **114**, 283-287 (2003).
- 685 69. Chiang, E.-P.I., Selhub, J., Bagley, P.J., Dallal, G. & Roubenoff, R. Pyridoxine supplementation
686 corrects vitamin B6 deficiency but does not improve inflammation in patients with rheumatoid
687 arthritis. *Arthritis research & therapy* **7**, R1404 (2005).
- 688 70. O'connor, Á. An overview of the role of diet in the treatment of rheumatoid arthritis. *Nutrition*
689 *bulletin* **39**, 74-88 (2014).
- 690 71. Selhub, J., *et al.* Dietary vitamin B6 intake modulates colonic inflammation in the IL10^{-/-} model
691 of inflammatory bowel disease. *The Journal of nutritional biochemistry* **24**, 2138-2143 (2013).
- 692 72. Zhang, X., *et al.* The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly
693 normalized after treatment. *Nature medicine* **21**, 895 (2015).
- 694 73. Bazin, T., *et al.* Microbiota Composition May Predict Anti-Tnf Alpha Response in
695 Spondyloarthritis Patients: an Exploratory Study. *Scientific reports* **8**, 5446 (2018).
- 696 74. Varela, E., *et al.* Colonisation by Faecalibacterium prausnitzii and maintenance of clinical
697 remission in patients with ulcerative colitis. *Alimentary pharmacology & therapeutics* **38**, 151-
698 161 (2013).
- 699 75. Maeda, Y., *et al.* Dysbiosis contributes to arthritis development via activation of autoreactive T
700 cells in the intestine. *Arthritis & rheumatology* **68**, 2646-2661 (2016).
- 701 76. Pianta, A., *et al.* Evidence of the immune relevance of Prevotella copri, a gut microbe, in patients
702 with rheumatoid arthritis. *Arthritis & rheumatology* **69**, 964-975 (2017).
- 703 77. Glick-Bauer, M. & Yeh, M.-C. The health advantage of a vegan diet: exploring the gut microbiota
704 connection. *Nutrients* **6**, 4822-4838 (2014).
- 705 78. Kovatcheva-Datchary, P., *et al.* Dietary fiber-induced improvement in glucose metabolism is
706 associated with increased abundance of Prevotella. *Cell metabolism* **22**, 971-982 (2015).
- 707 79. Kim, D. & Kim, W.U. can Prevotella copri be a causative pathobiont in rheumatoid arthritis?
708 *Arthritis & rheumatology* **68**, 2565-2567 (2016).
- 709 80. Wen, C., *et al.* Quantitative metagenomics reveals unique gut microbiome biomarkers in
710 ankylosing spondylitis. *Genome biology* **18**, 142 (2017).
- 711 81. Ley, R.E. Gut microbiota in 2015: Prevotella in the gut: choose carefully. *Nature Reviews*
712 *Gastroenterology & Hepatology* **13**, 69-70 (2016).
- 713 82. Ebringer, R., Cooke, D., Cawdell, D.R., Cowling, P. & Ebringer, A. Ankylosing spondylitis: klebsiella
714 and HL-A B27. *Rheumatology and rehabilitation* **16**, 190-196 (1977).
- 715 83. Stone, M., *et al.* Comparative immune responses to candidate arthritogenic bacteria do not
716 confirm a dominant role for Klebsiella pneumonia in the pathogenesis of familial ankylosing
717 spondylitis. *Rheumatology* **43**, 148-155 (2003).
- 718 84. Pugsley, A.P., Chapon, C. & Schwartz, M. Extracellular pullulanase of Klebsiella pneumoniae is a
719 lipoprotein. *Journal of bacteriology* **166**, 1083-1088 (1986).

- 720 85. Ebringer, A. & Wilson, C. The use of a low starch diet in the treatment of patients suffering from
721 ankylosing spondylitis. *Clinical rheumatology* **15**, 62-66 (1996).
- 722 86. Rashid, T., Wilson, C. & Ebringer, A. The link between ankylosing spondylitis, Crohn's disease,
723 Klebsiella, and starch consumption. *Clin Dev Immunol* **2013**, 872632 (2013).
- 724 87. Ciccia, F., *et al.* Dysbiosis and zonulin upregulation alter gut epithelial and vascular barriers in
725 patients with ankylosing spondylitis. *Ann Rheum Dis* **76**, 1123-1132 (2017).
- 726 88. Kubinak, J.L., *et al.* MHC variation sculpts individualized microbial communities that control
727 susceptibility to enteric infection. *Nature communications* **6**, 8642 (2015).
- 728 89. Dubois, P.C., *et al.* Multiple common variants for celiac disease influencing immune gene
729 expression. *Nature genetics* **42**, 295-302 (2010).
- 730 90. Brenner, O., *et al.* Loss of Runx3 function in leukocytes is associated with spontaneously
731 developed colitis and gastric mucosal hyperplasia. *Proceedings of the National Academy of*
732 *Sciences of the United States of America* **101**, 16016-16021 (2004).
- 733

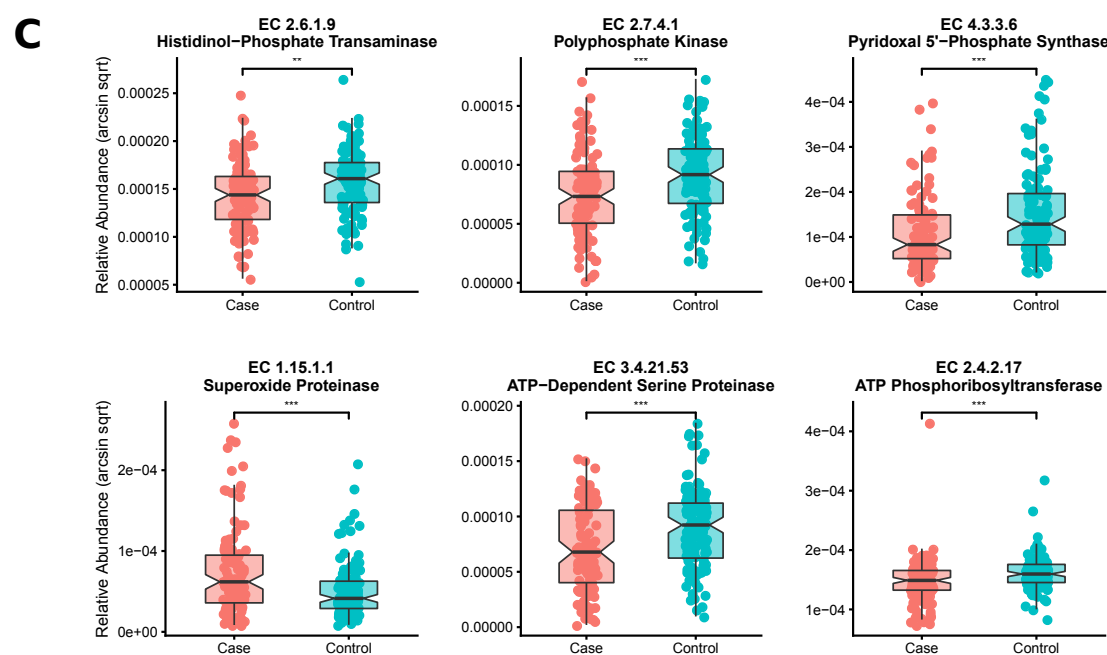
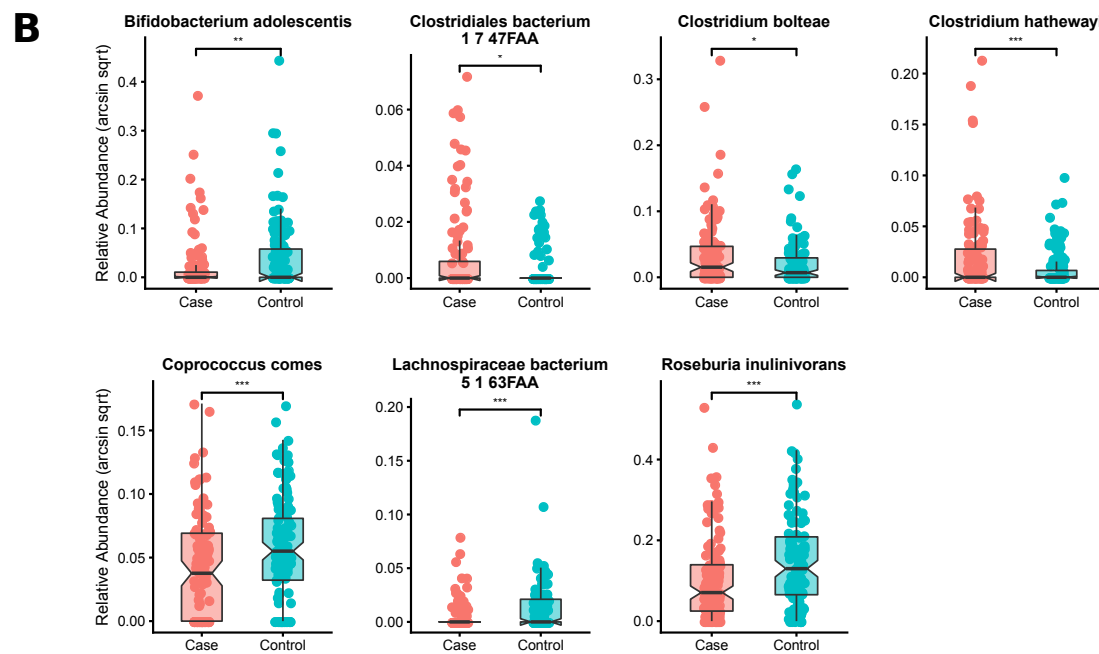
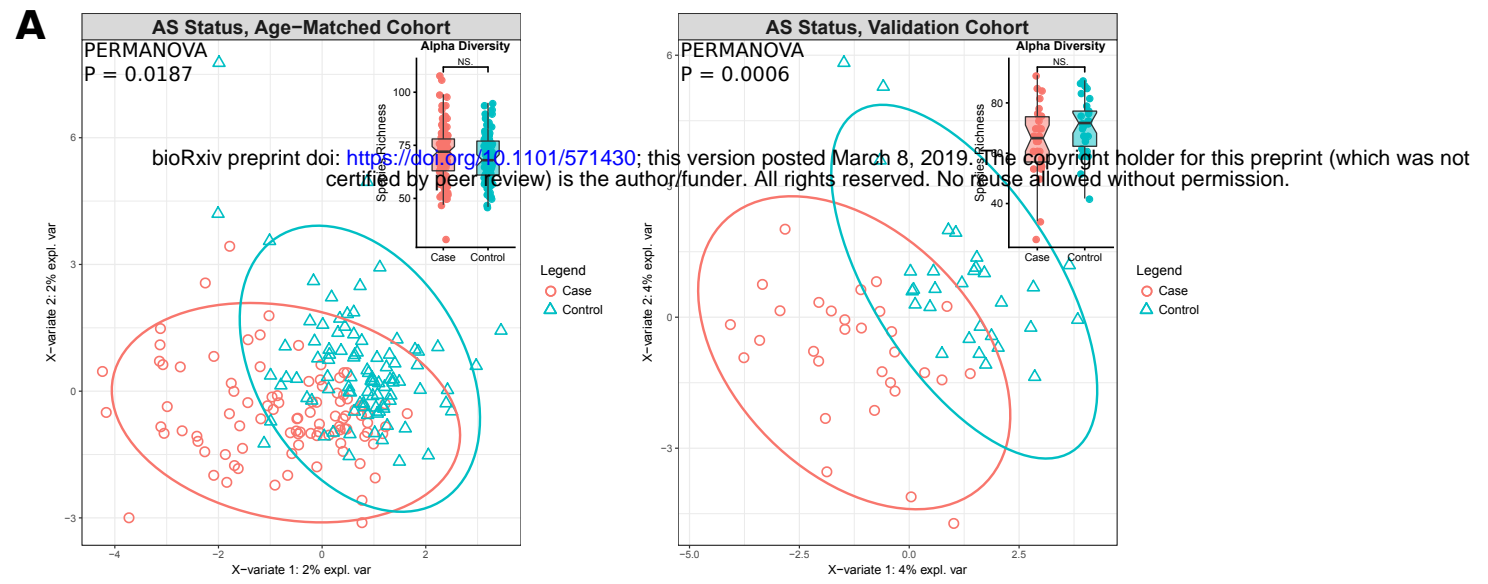


Figure 1: Taxonomic and functional dysbiosis observed in AS cases relative to healthy controls.
A. Alpha and beta diversity analysis. sPLSDA and PERMANOVA revealed community-level differences in taxonomic composition. **B.** Commonly-differentiated bacterial species from the discovery and validation cohorts. **C.** Commonly-differentiated KEGG Orthogroups from the discovery and validation cohorts. Bacterial species and KEGG Orthogroups exhibiting significant results according to multivariate linear modelling and Wilcoxon rank-sum tests are shown.

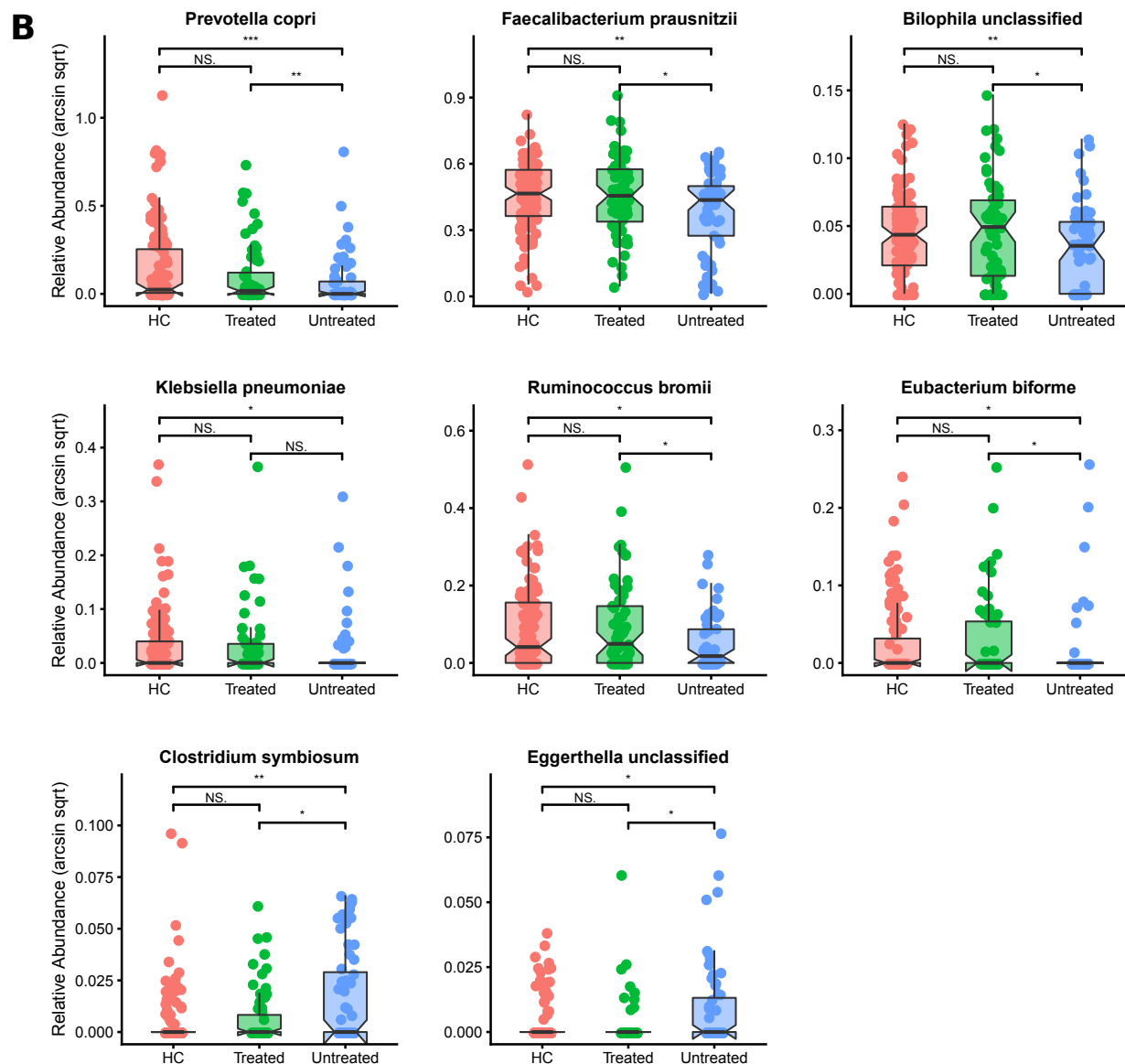
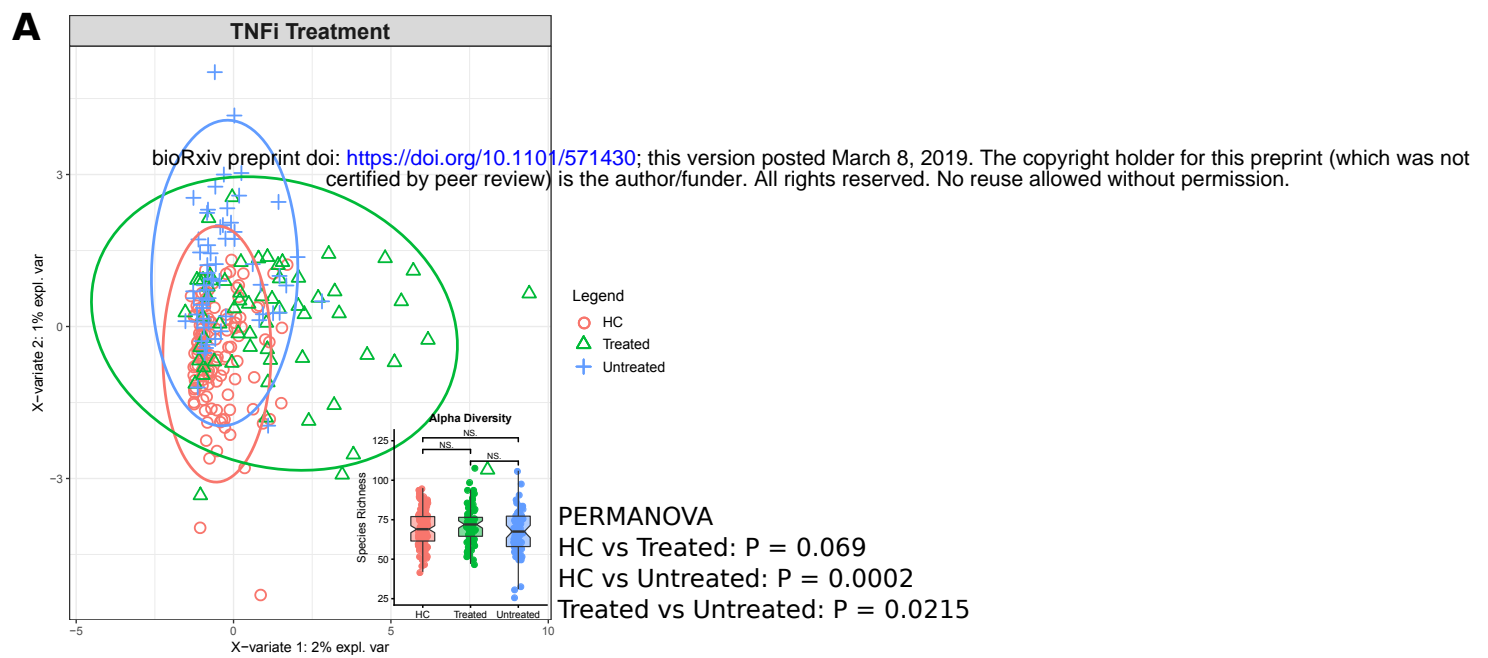


Figure 2: Effect of TNF α therapy upon the microbiome. **A.** Alpha and beta diversity analysis. sPLSDA and PERMANOVA revealed community-level differences in taxonomic composition. **B.** Bacterial species modulated by the effects of TNF α treatment. Bacterial species exhibiting significant results according to multivariate linear modelling and Wilcoxon rank-sum tests are shown.

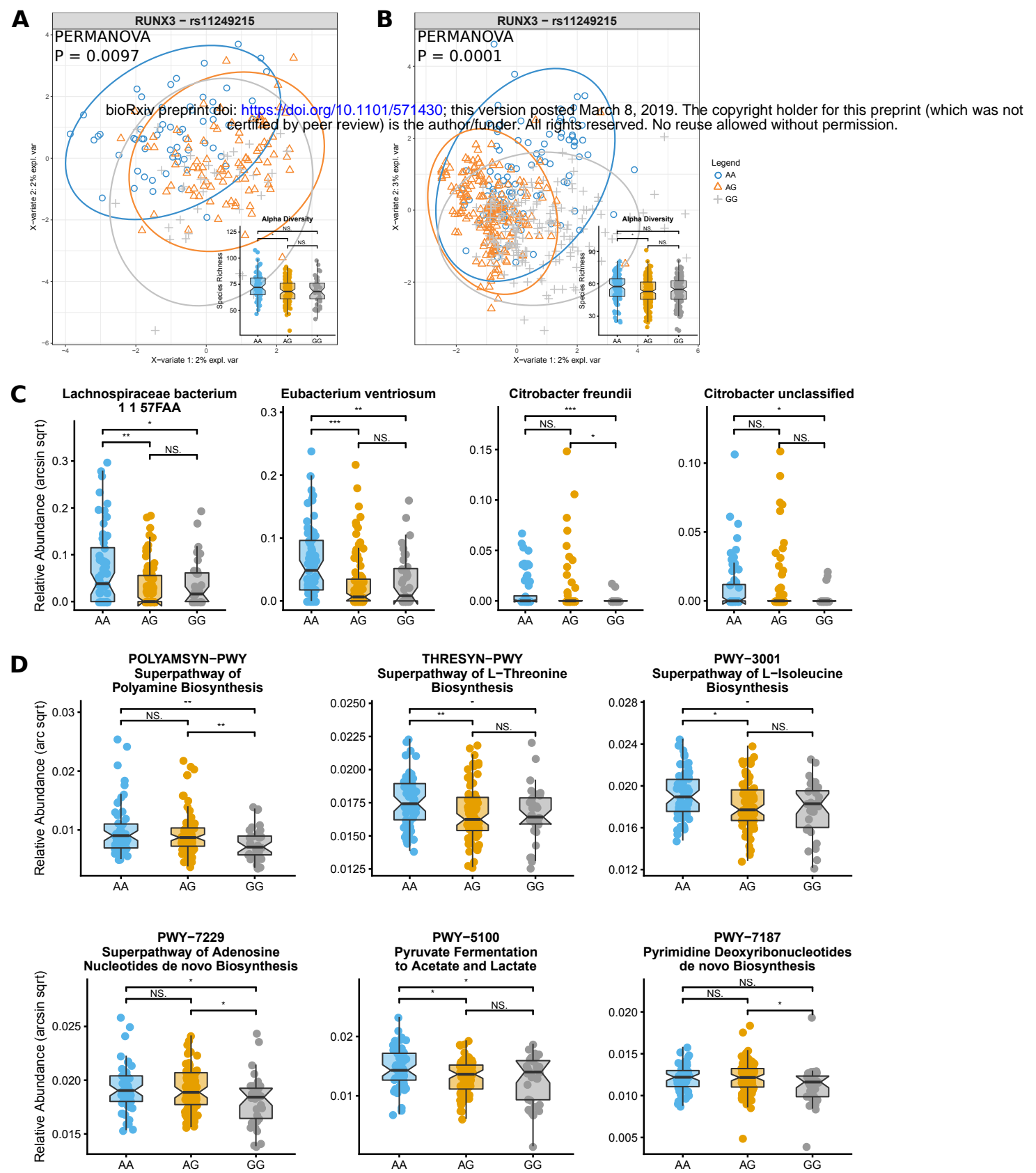


Figure 3: Effect of RUNX3 variants upon the microbiome. **A.** sPLSDA, alpha diversity and PERMANOVA community-level taxonomic analysis of the current study. **B.** sPLSDA, alpha diversity and PERMANOVA community-level taxonomic analysis of a recent 16S metabarcode study of healthy individuals. **C.** Modulated bacterial species according to significant results from multivariate linear modelling and Wilcoxon rank-sum testing. **D.** Modulated MetaCyc metabolic pathways, according to significant results from multivariate linear modelling and Wilcoxon rank-sum testing.

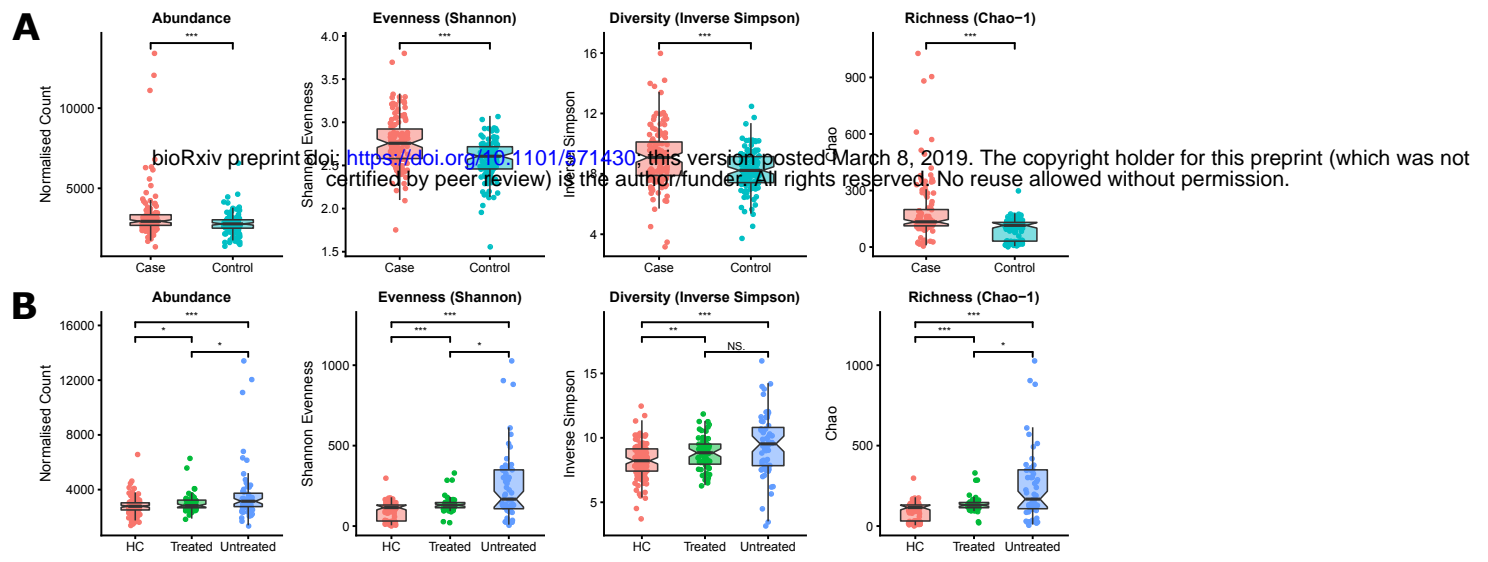


Figure 4: A. Enrichment, both in terms of abundance and diversity, of bacterial peptides homologous to HLA-B27 epitopes in AS cases relative to healthy controls. **B.** Differential abundance and diversity of bacterial peptides homologous to HLA-B27 epitopes in TNFi-treated and –untreated cases, and healthy controls.

Epitope ID	HLA-B27 subtype	Sequence
490238	B2702;B2704	ARFKSNVTKTMKGFY
493378	B2704;B2705;B2709	MRLPAQLLGLLM
493505	B2702	NRHYTFYVW
491582	B2702	GRINPNSGCTNY
494083	B2702	QTTFLVDNKKVFGTHL
494108	B2702	RQIMTGFGEYSY
434944	B2702;B2705;B2707;B2709	ARTPHWALF
492935	B2703	KRWESERVLSF
493023	B2705	LPVNLSTSGPF
492689	B2703;B2707;B2709	KRFDDKYTLKLT
492876	B2702	KRNEDESPNKLY
442808	B2702;B2703;B2708	ARLDIDPETITW
494092	B2702	RKFQPYKPFYY
495095	B2708	SRLEQGEEPWWL
490897	B2702	ERIAEFNQLQF
493564	B2702	NRQIVSGSRDKTIKLW
445935	B2702;B2703;B2707;B2708;B2709	KRNTFVGTPFWM
491482	B2702;B2703;B2704;B2705;B2706;B2707;B2708	GRFTIKSDVWSF
490741	B2702	ATTAALLLEAQAATGFLVDPVR
492710	B2703	KRFFFDVGSNKY
494342	B2707;B2709	RRIMRPTDVPDQGL
447192	B2702	QRGLWGGEW
492970	B2702	KRYDEVEAEGY
493651	B2702	QRAIQVDPNYAY
491073	B2702	FQWMSSRVSPNTLW
494077	B2705	QRYSLLPFWY

Table 1: Bacterial peptides homologous to HLA-B27-presented epitopes which were commonly enriched in the discovery and validation cohorts for AS cases.

bioRxiv preprint doi: <https://doi.org/10.1101/571430>; this version posted March 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.