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

Jian Yin[1*], Peter R. Sternes[2*], Mingbang Wang[3], Mark Morrison[4], Jing Song[1], Ting Li[1], Ling Zhou[1], Xin Wu[1], Fusheng He[5], Jian Zhu[6], Matthew A. Brown[2 º], Huji Xu[1,6,7º].

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

*These authors contributed equally to this manuscript.

º These authors contributed equally to this manuscript.

Corresponding Author:
Professor Matthew A. Brown,
Institute of Health and Biomedical Innovation,
Queensland University of Technology,
Translational Research Institute,
Princess Alexandra Hospital,
Woolloongabba, Brisbane,
Australia.
Telephone: +61 7 3443 7366
Fax : +61 7 3443 7102
matt.brown@qut.edu.au

Corresponding Author:
Professor Huji Xu
Beijing Tsinghua Changgeng Hospital
School of Clinical Medicine
Tsinghua University, Beijing 100084, China.
Phone: +86-10-6279 2998
Fax: +86-10-6278 9513
email: huji_xu@tsinghua.edu.cn

**ABSTRACT**

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

## INTRODUCTION

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

AS has been shown in both twin and unrelated case/control studies to be highly heritable (twins >90% heritability [4,5], unrelated case/control common variant heritability 69% (http://www.nealelab.is/blog/2017/9/15/heritability-of-2000-traits-and-disorders-in-the-uk-biobank)). Over past decade, at least 116 susceptibility genes have been identified, contributing 29% of the overall risk of the disease [6]. There is substantial evidence suggesting that the interaction between host genetics and gut microbiome is a key driver of the pathogenesis of AS. The high disease heritability indicates that the environmental factors involved in the disease are likely to be ubiquitous. Reactive arthritis is a spondyloarthritis sharing many clinical and genetic features with AS, and is known to be caused by bacterial infections of the gut or urinary tract; a subset of reactive arthritis patients go on to develop AS. About 60% AS patients suffered from subclinical bowel inflammation and 10% of them can be diagnosed as inflammatory bowel disease (IBD) [7]. There is considerable overlap in the overall heritability of AS and IBD [8], the two diseases are often co-familial [9], and many shared genetic associations have been identified [10]. A bioinformatic study showed that AS susceptibility genes specifically enriched in gut cells

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

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

## MATERIALS AND METHODS

### Subject recruitment

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

### DNA microarray and subject genotyping

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

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

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

**Shotgun metagenome sequencing**

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

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

Manual inspection and QC of sequencing reads was conducted using FastQC v10.1 [26]. Paired-end reads were joined using PEAR v0.9.10 [27] and adapters were trimmed using Trimmomatic v0.36 [28]. Contaminant sequences, such as those mapping to human or PhiX genomes, were filtered using Bowtie2 v2.3.4 [29] and the remaining reads were counted and subsampled to an equal sequencing depth of 3,520,000 sequencing reads per sample using SeqTK v1.0 [30]. MetaPhlAn2 v2.6.0 [31] was used for taxonomic classification, PanPhlAn v1.2.2 [32] was used for strain-level profiling utilising pre-computed pan-genome

154 references where possible, and HUMAnN2 v0.11.1 [33] 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-formatted [34], 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 study [37] 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 v6.3.1 [38], 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 v2.4-5 [39]. 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 v0.0.5 [40] and Wilcoxon rank-sum tests as implemented in R [41]. 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,*

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

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

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

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

**Effect of TNFi therapy upon the microbiome**

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

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

autoimmune diseases [42]. TNFi therapy appeared to partially normalise the dysbiotic bacterial species and KEGG Orthogroups observed in AS cases relative to healthy controls shown in Figures 1B and 1C, however no statistically significant differences between treated and untreated cases were observed, potentially due to sample size constraints (Supplementary Figure 4).

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

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

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

**Effect of host genotype upon the microbiome**

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

Considering non-MHC loci, an association was noted for rs11249215, a SNP in the promoter of runt-related transcription factor 3 (*RUNX3*) gene known to be associated with AS [45,46]. This variant was associated with a shift in the microbiome of both AS cases and healthy controls (combined P = 0.0097). Furthermore, sPLSDA revealed that the degree alteration appears dependent on whether the host carried a heterozygous or homozygous genotype (Figure 3A), with the homozygous genotype resulting in a more substantial shift. As further confirmation, we analysed a recently published 16S metabarcoding dataset [47] of 107 healthy control subjects which were sampled from six different body sites. This analysis re-confirmed discrimination of the microbiomes based on genotype (PERMANOVA; P = 0.0001) (Figure 3B). The *RUNX3* SNP had no observable effect upon the dysbiotic bacterial species and KEGG Orthogroups outlined in Figures 1B and 1C, however its effects upon species richness and community composition (Figure 3A) suggest that further research is required to confirm a role in AS pathogenesis 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

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

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

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

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

To investigate the abundance of bacterial peptides homologous to HLA-B27 epitopes in AS cases and healthy controls, translated nucleotide searches were performed against IEBD v3.0, annotated according to a published study [37] and counted. Significant enrichment of these peptide sequences was observed in 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.

## DISCUSSION

### Gut dysbiosis in ankylosing spondylitis

This study re-confirmed the occurrence of bacterial gut dysbiosis in AS cases and identified seven bacterial species which were commonly differentiated between cases and controls in both the discovery and validation cohorts (Figure 1B). Two of these species, *Bifidobacterium adolescentis* and *Coprococcus comes*, have been noted for their depletion in Crohn's Disease [51] and were also observed to be depleted in AS cases in this study. An additional two species previously reported to be associated with AS, *Prevotella copri* and *Dialister invisus*, were found to be differentially abundant in either the discovery or validation cohorts (Supplementary Table 2). In the case of *Prevotella copri*, previous studies have demonstrated enrichment in new onset RA cases yet depletion in chronic RA cases [52]. Consistent with these findings, our study found that *Prevotella copri* was depleted in AS cases within the non-age-matched cohort, for which the demographics were heavily skewed towards older AS patients with long-standing disease (Supplementary Table 1). Previous studies in AS have shown increases in *Prevotellaceae* [14], or specifically with this species [15]. As discussed below, *Prevotella copri* carriage normalised with TNFi treatment. Further studies will be required to determine if *Prevotella copri* carriage changes with disease duration, as has been reported in RA.

Carriage of *Dialister* species has been previously associated with disease activity in spondylarthrosis patients [17], but the carriage of *Dialister invisus* has been reported to be decreased in inflammatory bowel disease (IBD) [53,54]. Whilst we found enrichment of *Dialister invisus* in AS cases in the discovery cohort, this was not confirmed in the validation cohort, nor has it been reported in other AS studies. Its pathogenic significance is therefore uncertain.

379 Of particularly 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 IBD [54-59], and has been previously shown to be depleted in the disease enthesitis-related
382 arthritis, a paediatric disease-classification which includes children with ankylosing spondylitis [60].
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 differentiation [61], 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 patients [66-71]. 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 party restored following treatment [72]. 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 size [73]. 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 study [55]
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
424 [74]. Another study demonstrated that treatment of infliximab completely restored *Faecalibacterium*
425 *prausnitzii* concentrations from zero to $1.4 \times 10^{10}$ bacteria/mL within few days [52].
426

Another important microbe, *Prevotella copri*, was observed to be enriched to levels closely matched to that of healthy controls following TNFi treatment. Abundance of *Prevotella copri* has previously been shown to be enriched in untreated new onset RA patients yet depleted in chronic RA cases, patients with psoriatic arthritis and healthy controls [52]. Colonisation of SKG mice with *Prevotella copri*-dominated microbiota from RA patients exhibited increased number of Th17 cells in the large intestine [75]. HLA-DR-presented peptides (T cell epitopes) from *Prevotella copri* were recently found to stimulate Th1 responses in 42% of new onset RA cases, with subgroups of RA patients demonstrating differential IgG or IgA immune reactivity, providing evidence that *Prevotella copri* is immune-relevant in RA pathogenesis [76]. Additionally, the presence of the *HLA-DRB1* risk allele, which influences disease severity, in RA patients was found to be inversely correlated with *Prevotella copri* abundance [52,77-79]. A recent study of Chinese AS patients revealed enrichment of *Prevotella copri*, as well as *Prevotella melaninogenica* and *Prevotella* sp. C561 [15,80]. Contrasting these results, in the current study we observed depletion of *Prevotella copri* in untreated AS cases, which was restored to the healthy control levels in TNFi-treated patients. These seemingly conflicting reports of *Prevotella copri* abundance may be explained by the large degree of intraspecific genetic diversity of *Prevotella copri* strains, with strain variation adding an additional layer of complexity for predicting the function of *Prevotella copri* in the gut. The *Prevotella* genus also contains members that may be beneficial, and which do not function as pathobionts [77-79], with observed enrichment in healthy individuals. Taken together, our results which demonstrate a modulation of *Prevotella copri* abundance in TNFi-treated cases is a noteworthy observation, however without a stronger grasp of the strain-level genome variation within this taxon and their prevalence across our cohort, attempts to therapeutically modulate and predict the effects of *Prevotella copri* remains a significant challenge [81].

450 *Klebsiella pneumoniae* has also been suggested to play a significant role in AS pathogenesis [82], although
451 this remains controversial [83]. *Klebsiella pneumoniae* notably produces pullulanase, a starch-debranching
452 enzyme which enables the degradation of starch into simple sugars [84]. 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 patients [86], 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 ($P = 0.014$; $R^2 = -0.043$) (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.

**Enrichment of potentially arthritogenic bacterial peptides**

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

**Effect of host genotype upon the microbiome**

Very recently, it was demonstrated that HLA-B27 is associated with a significant shift in the microbiome in healthy individuals [47]. Furthermore, in mouse models, MHC polymorphisms were demonstrated to contribute to an individual's microbial composition, thus influencing health [88]. Our investigation revealed an additional AS-associated SNP, rs11249215 in *RUNX3* [45,46], which was also correlated with a significant shift in bacterial composition (Figure 3A). This result was replicated in a confirmatory dataset of healthy Caucasian individuals (Figure 3B). In addition to potential roles in autoimmune diseases, 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

## CONCLUSION

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

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGEMENTS

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

## REFERENCES


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

20. Linden, S.V.D., Valkenburg, H.A. & Cats, A. Evaluation of diagnostic criteria for ankylosing spondylitis. *Arthritis & rheumatology* **27**, 361-368 (1984).
21. Garrett, S.*, et al.* A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. *J Rheumatol* **21**, 2286-2291 (1994).
22. Calin, A.*, et al.* A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. *J Rheumatol* **21**, 2281-2285 (1994).
23. Das, S.*, et al.* Next-generation genotype imputation service and methods. *Nature genetics* **48**, 1284 (2016).
24. Jia, X.*, et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PloS one* **8**, e64683 (2013).
25. Rentería, M.E., Cortes, A. & Medland, S.E. Using PLINK for genome-wide association studies (GWAS) and data analysis. in *Genome-Wide Association Studies and Genomic Prediction* 193-213 (Springer, 2013).
26. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).
27. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620 (2013).
28. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
29. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357 (2012).
30. Li, H. seqtk Toolkit for processing sequences in FASTA/Q formats. (GitHub, 2012).
31. Truong, D.T.*, et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature methods* **12**, 902 (2015).
32. Scholz, M.*, et al.* Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nature methods* **13**, 435 (2016).
33. Abubucker, S.*, et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS computational biology* **8**, e1002358 (2012).
34. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).
35. Vita, R.*, et al.* The immune epitope database (IEDB) 3.0. *Nucleic acids research* **43**, D405-D412 (2014).
36. Kim, Y.*, et al.* Immune epitope database analysis resource. *Nucleic acids research* **40**, W525-W530 (2012).
37. Schittenhelm, R.B., Sian, T.C., Wilmann, P.G., Dudek, N.L. & Purcell, A.W. Revisiting the arthritogenic peptide theory: quantitative not qualitative changes in the peptide repertoire of HLA–B27 allotypes. *Arthritis & rheumatology* **67**, 702-713 (2015).
38. Lê Cao, K.-A.*, et al.* MixMC: a multivariate statistical framework to gain insight into microbial communities. *PloS one* **11**, e0160169 (2016).
39. Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927-930 (2003).
40. Morgan, X.C.*, et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome biology* **13**, R79 (2012).
41. Ihaka, R. & Gentleman, R. R: a language for data analysis and graphics. *Journal of computational and graphical statistics* **5**, 299-314 (1996).
42. Forbes, J.D., Van Domselaar, G. & Bernstein, C.N. The gut microbiota in immune-mediated inflammatory diseases. *Frontiers in microbiology* **7**, 1081 (2016).

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

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

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

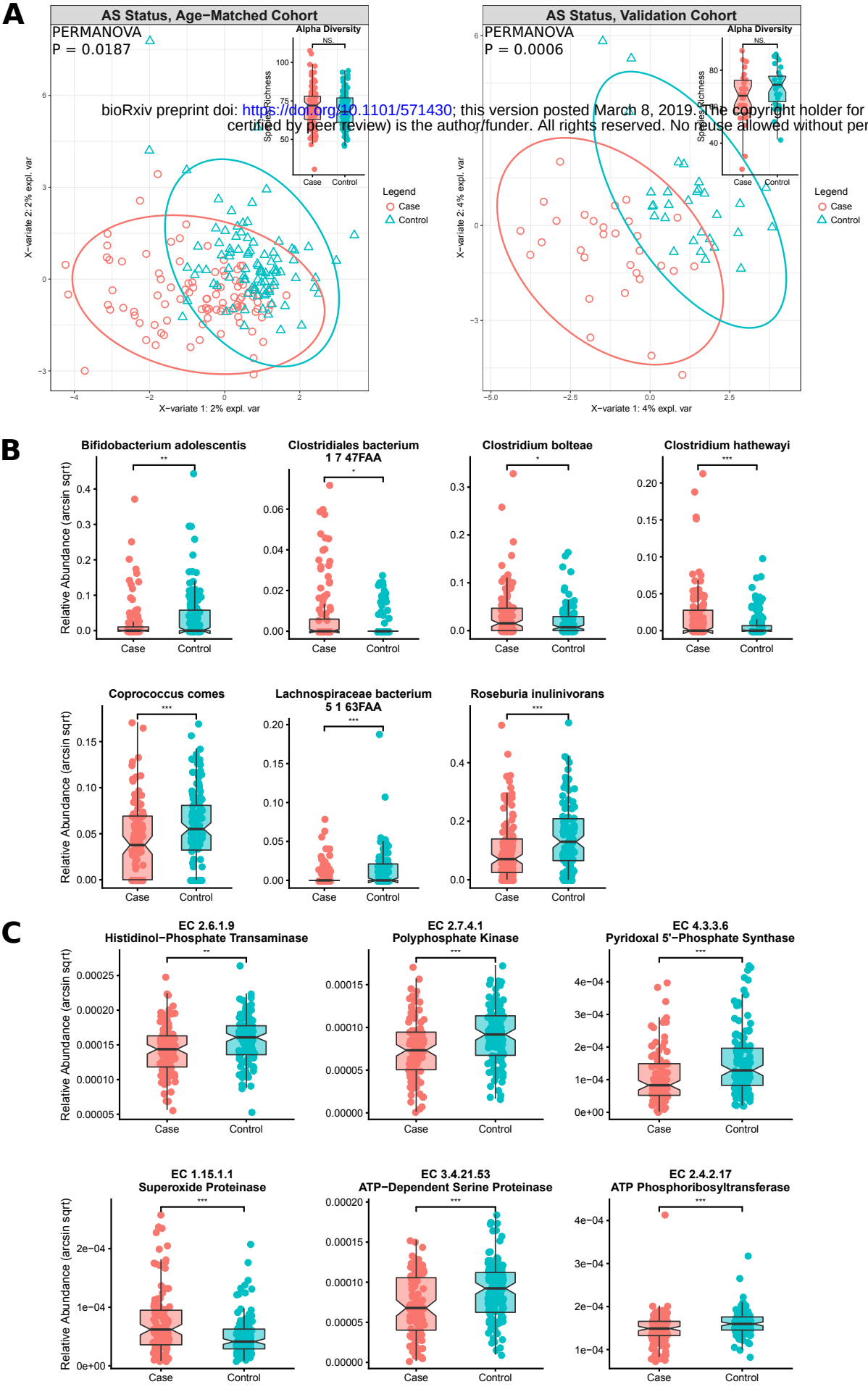
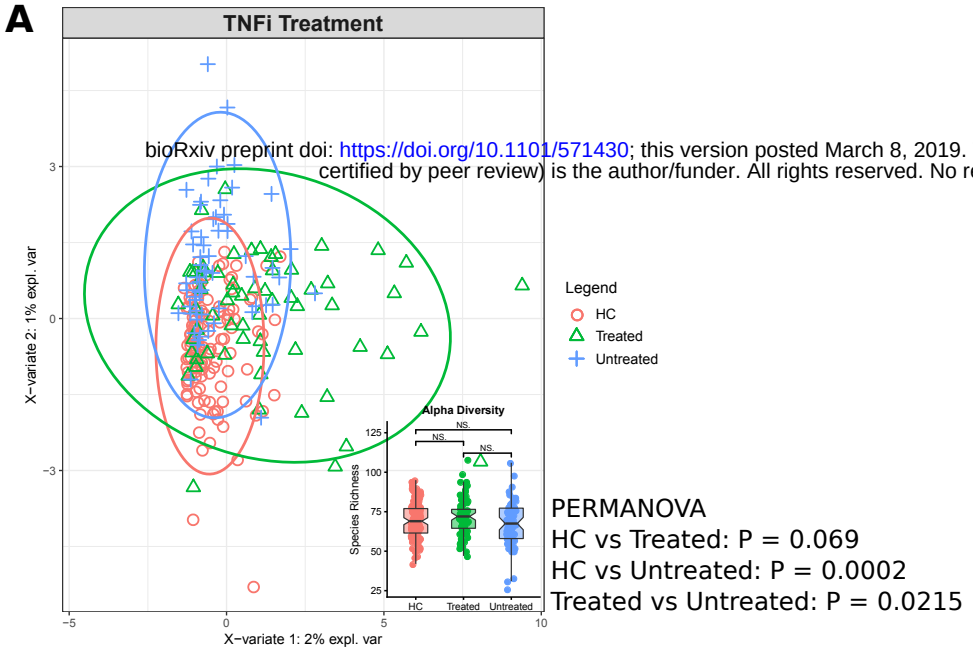



**Figure 1:** Taxonomic and functional dysbiosis observed in AS cases relative to healthy controls. **A.** Alpha and beta diversity analysis. sPLSDA and PERMANOVA revealed community-level differences in taxonomic composition. **B.** Commonly-differentiated bacterial species from the discovery and validation 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.



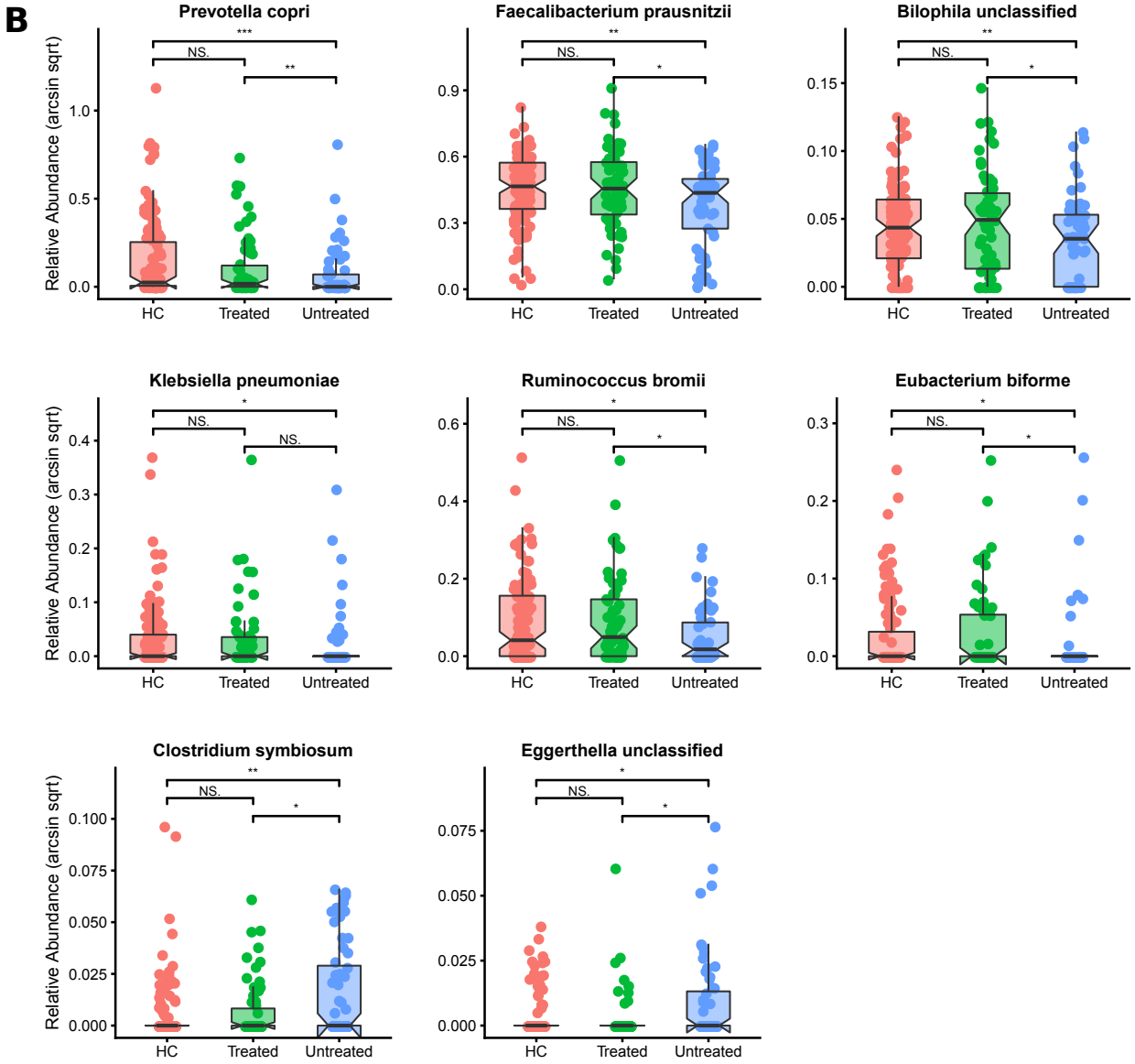




**Figure 2:** Effect of TNFi therapy upon the microbiome. **A.** Alpha and beta diversity analysis. sPLSDA and PERMANOVA revealed community-level differences in taxonomic composition. **B.** Bacterial species modulated by the effects of TNFi treatment. Bacterial species exhibiting significant results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

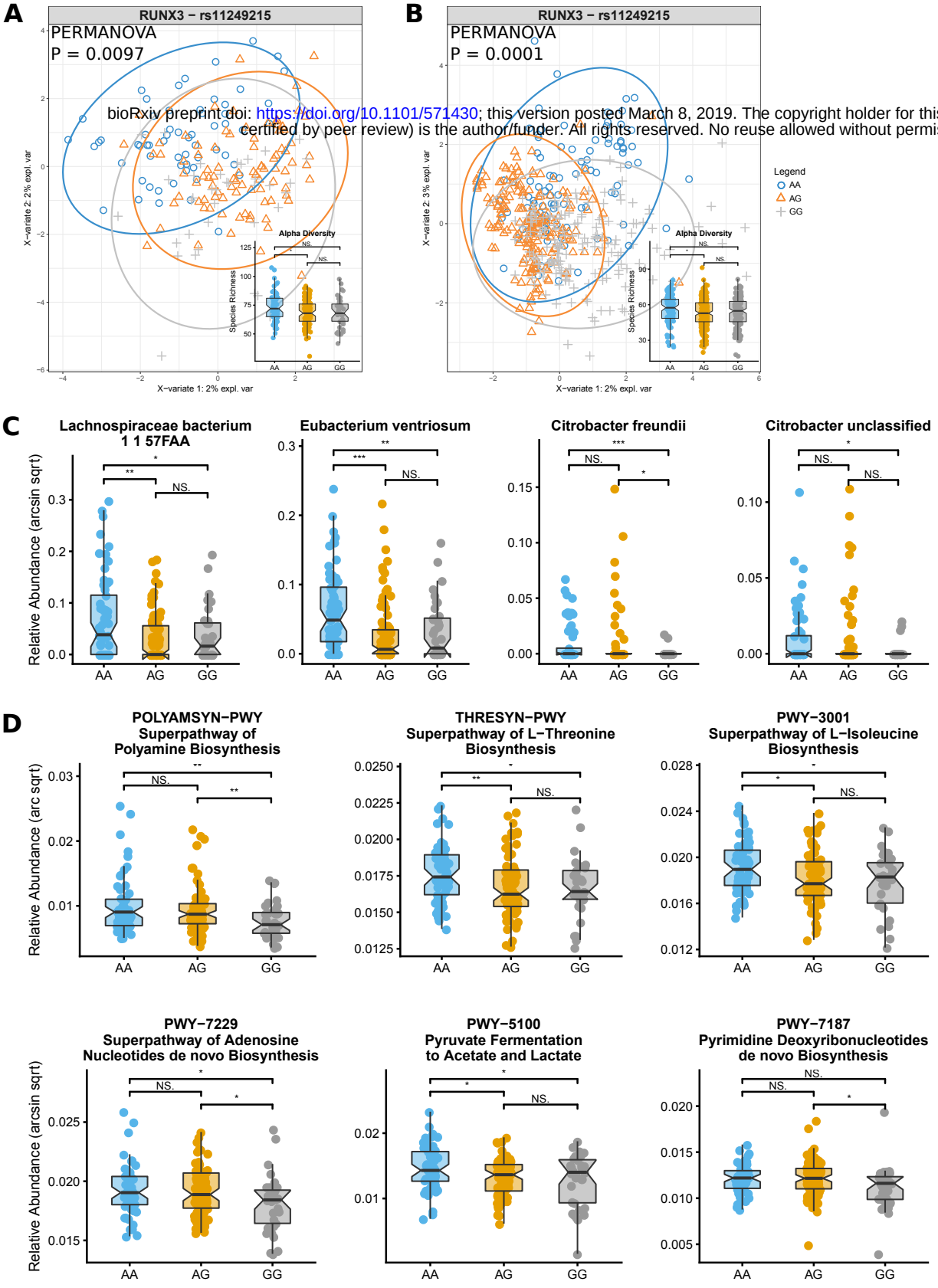



**Figure 3:** Effect of RUNX3 variants upon the microbiome. **A.** sPLSDA, alpha diversity and PERMANOVA community-level taxonomic analysis of the current study. **B.** sPSLDA, alpha diversity and PERMANOVA community-level taxonomic analysis of a recent 16S metabarcode study of healthy individuals. **C.** 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.

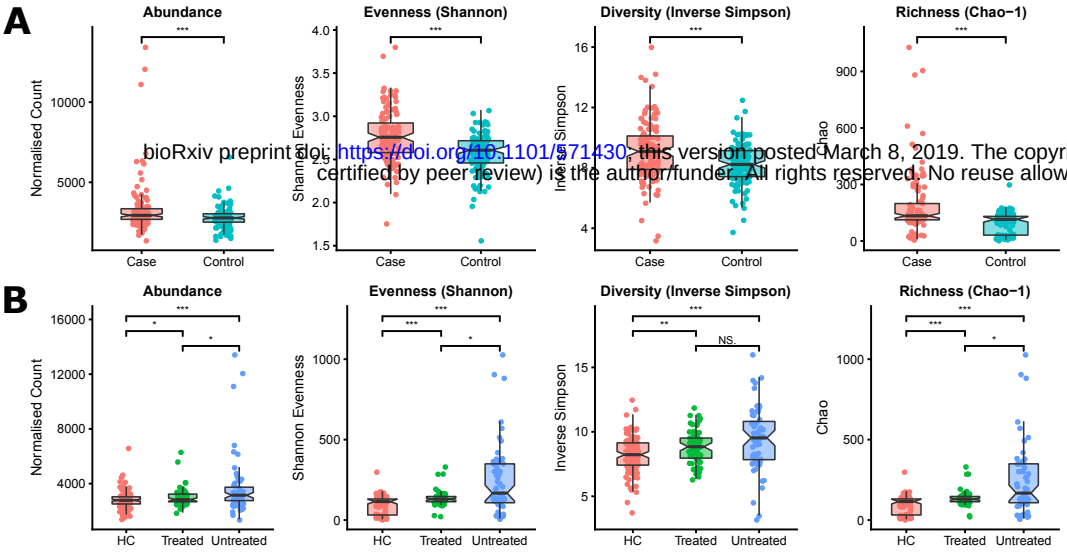



**Figure 4: A.** Enrichment, both in terms of abundance and diversity, of bacterial peptides homologous to HLA-B27 epitopes in AS cases relative to healthy controls. **B.** Differential abundance and diversity of bacterial peptides homologous to HLA-B27 epitopes in TNFi-treated and –untreated cases,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

| Epitope ID | HLA−B27 subtype | Sequence |
|---|---|---|
| 490238 | B2702;B2704 | ARFKSNVTKTMKGFEY |
| 493378 | B2704;B2705;B2709 | MRLPAQLLGLLM |
| 493505 | B2702 | NRHYTFYVW |
| 491582 | B2702 | GRNPNSGSTNY |
| 494083 | B2702 | QTTFLVDNKKVFGTHL |
| 494108 | B2702 | RQIMTGFGELSY |
| 434944 | B2702;B2705;B2707;B2709 | ARTPHWALF |
| 492935 | B2703 | KRWESERVLSF |
| 493023 | B2705 | LPVNLLLSTSGPF |
| 492689 | B2703;B2707;B2709 | KRFDDKYTLKLT |
| 492876 | B2702 | KRNEDEDSPNKLY |
| 442808 | B2702;B2703;B2708 | ARLDIDPETITW |
| 494092 | B2702 | RKFQPYKPFYY |
| 495095 | B2708 | SRLEQGEEPWVL |
| 490897 | B2702 | ERIATEFNQLQF |
| 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 | QRGLWGGEGW |
| 492970 | B2702 | KRYYDEVEAEGY |
| 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.