medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Plasma Metabolomic Profiling in Patients with Rheumatoid Arthritis Identifies Biochemical Features Predictive of Quantitative Disease Activity

Benjamin Hur[1,2], Vinod K. Gupta[1,2], Harvey Huang[3], Kerry A. Wright[4], Kenneth J. Warrington[4], Veena Taneja[5], John M. Davis III[4,*], and Jaeyun Sung[1,2,4,*, †]

[1]Microbiome Program, Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA
[2]Division of Surgery Research, Department of Surgery, Mayo Clinic, Rochester, MN, USA
[3]Mayo Clinic Medical Scientist Training Program, Mayo Clinic, Rochester, MN, USA
[4]Division of Rheumatology, Department of Medicine, Mayo Clinic, Rochester, MN, USA
[5]Department of Immunology, Mayo Clinic, Rochester, MN, USA
[*]Contributed equally as co-senior authors
[†]Corresponding author: Sung.Jaeyun@mayo.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background: Rheumatoid arthritis (RA) is a chronic, autoimmune disorder characterized by joint inflammation and pain. In patients with RA, metabolomic approaches, i.e., high-throughput profiling of small-molecule metabolites, on plasma or serum has thus far enabled the discovery of biomarkers for clinical subgroups, risk factors, and predictors of treatment response. Despite these recent advancements, the identification of blood metabolites that reflect quantitative disease activity remains an important challenge in precision medicine for RA. Herein, we use global plasma metabolomic profiling analyses to detect metabolites associated with, and predictive of, quantitative disease activity in patients with RA.

Methods: Ultra-high performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) was performed on a discovery cohort consisting of 128 plasma samples from 64 RA patients, and on a validation cohort of 12 samples from 12 patients. The resulting metabolomic profiles were analyzed with two different strategies to find metabolites associated with RA disease activity defined by the Disease Activity Score-28 using C-reactive protein (DAS28-CRP). More specifically, mixed-effects regression models were used to identify metabolites differentially abundant between two disease activity groups ('lower', DAS28-CRP $\leq$ 3.2; and 'higher', DAS28-CRP > 3.2); and to identify metabolites significantly associated with DAS28-CRP scores. A generalized linear model (GLM) was then constructed for estimating DAS28-CRP using plasma metabolite abundances. Finally, for associating metabolites with CRP (an indicator of inflammation), metabolites differentially abundant between two patient groups ('low-CRP', CRP $\leq$ 3.0 mg/L; 'high-CRP', CRP > 3.0 mg/L) were investigated.

Results: We identified 33 metabolites differentially abundant between lower and higher disease activity groups ($P < 0.05$). Additionally, we identified 51 metabolites associated with DAS28-CRP ($P < 0.05$). A GLM based upon these 51 metabolites resulted in higher prediction accuracy (mean absolute error [MAE]±SD: 1.51±1.77) compared to a GLM without feature selection (MAE±SD: 2.02±2.21). The predictive value of this feature set was further demonstrated on a validation cohort of twelve plasma samples, wherein we observed a stronger correlation between predicted vs. actual DAS28-CRP (with feature selection: Spearman's ρ = 0.69, 95% CI: [0.18, 0.90]; without feature selection: Spearman's ρ = 0.18, 95% CI: [-0.44, 0.68]). Lastly, among all identified metabolites, the abundances of eight were significantly associated with CRP patient groups while controlling for potential confounders ($P < 0.05$).

Conclusions: We demonstrate for the first time the prediction of quantitative disease activity in RA using plasma metabolomes. The metabolites identified herein provide insight into circulating pro-/anti-inflammatory metabolic signatures that reflect disease activity and inflammatory status in RA patients.

Keywords: rheumatoid arthritis, metabolomics, plasma metabolites, DAS28-CRP, biomarker, machine-learning, inflammation

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Background

Rheumatoid arthritis (RA) is a chronic, autoimmune inflammatory disease primarily affecting the small diarthrodial joints and other organ systems [1-4] that can eventually lead to bone/cartilage erosion, joint deformity, loss in mobility, and organ damage [5]. Known to be associated with a variety of factors, such as genetic susceptibility [6], age [7], sex [8], smoking status [9], and dietary habits [10], RA is diagnosed in nearly 5 per 1000 adults worldwide, and women are 2 to 3 times more likely to develop RA than men [5]. In addition, as is the case with many complex and progressive disorders, patients with RA exhibit vast heterogeneity in clinical symptoms (e.g., joint inflammation, swelling, pain, stiffness) [11], and in responses to methotrexate and other disease-modifying anti-rheumatic drugs (DMARDs) [12]. Furthermore, immune cells (mainly, B-cells, T-cells, and macrophages) and cytokines are known to be implicated in RA pathogenesis [13]. For example, Haringman *et al.* observed that the abundance of macrophages in synovial tissue was positively correlated with disease activity [14], and Chung *et al.* identified significant differences in levels of multiple cytokines (e.g., IL-6, IL-11, LIF) between RA and healthy controls [15]. In this regard, further understanding of the pathophysiological mechanisms that drive either progression or remission in RA disease activity would be important for identifying prognostic factors and developing more effective treatments [5, 16].

Having practical measures of disease activity is essential for determining the course of RA treatment and for monitoring patient response [3]. To this end, several studies have suggested strategies to quantify (or categorize) RA disease activity by using clinical and inflammatory core components, which include, but are not limited to, the number of tender and swollen joints, erythrocyte sedimentation rate (ESR), serum C-reactive protein (CRP) levels, and patients' pain levels [17-20]. Among these various strategies, the modified Disease Activity Score that considers 28 joints (DAS28) with either ESR (DAS28-ESR) or CRP (DAS28-CRP) is currently one of the most well-recognized and recommended measures in RA [20].

An emerging area of RA research is in using high-throughput metabolomic profiling approaches, which comprehensively measure all small-molecule biochemicals in a biological specimen (e.g., plasma, serum, urine, synovial fluid, etc.) to enable biomarker discovery and novel insights into the biochemical processes governing disease pathophysiology [11, 21-23]. In particular, recent studies have demonstrated the promise of using such metabolomic technologies on patient-derived biospecimens for classifying patients with RA according to their disease activity categories [21, 24, 25], and for identifying metabolic signatures predictive of treatment response [26-29]. For instance, Teitsma *et al.* used metabolomic profiling in serum samples from early RA patients to identify metabolites and metabolic pathways that were significantly associated with sustained, drug-free remission (DAS28 < 2.6) after tocilizumab- or methotrexate-based therapy [24]. Likewise, Sasaki *et al.* identified 15 and 20 metabolites in plasma and urine, respectively, that were differentially abundant between active RA (DAS28-ESR $\geq$ 3.2) and inactive RA (DAS28-ESR < 3.2) [25]. These findings suggest that a wider application of global metabolomic profiling—coupled with advanced analytics [30]—can lead to the discovery of novel and predictive biomarkers that complement current standard laboratory tests for assessing disease activity in RA.

To date, a global metabolomic profiling analysis to demonstrate the predictive value of blood biochemicals in estimating disease activity scores for patients with RA, has remained elusive. In this study, on 128 plasma metabolomic profiles from 64 RA patients, we utilize a multi-approach analysis to uncover metabolites that reflect and predict RA disease activity. First, we identify metabolites that stratify patients of 'higher' (DAS28-CRP $\geq$ 3.2) and 'lower' (DAS28-CRP < 3.2) disease activity groups. Next, we pinpoint

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

specific metabolites that significantly associate with DAS28-CRP. Interestingly, a few of the metabolites identified through these two approaches were able to differentiate between two groups of patients divided according to their C-reactive protein (CRP) levels in blood ('high-CRP', CRP > 3.0 mg/L; 'low-CRP', CRP ≤ 3.0 mg/L); these metabolites may possibly reflect metabolic perturbations affected by worsening inflammatory activity. Finally, we utilize a machine-learning technique to predict DAS28-CRP with plasma metabolite abundances. Importantly, we find that the feature selection step led to improved performance in predicting quantitative disease activity, and this translated reasonably well to a validation cohort. Taken together, our findings described herein support a key role for high-throughput metabolomic technologies in identifying blood-borne metabolic signatures of RA disease activity, and lay the groundwork for monitoring disease progression and systemic inflammation using blood samples alone.

Materials and Methods

Study Population, Subject Enrollment, Sample Collection, and Demographic Characteristics

The study population consisted of consecutive patients with RA attending the outpatient practice of the Division of Rheumatology at Mayo Clinic in Rochester, Minnesota. Eligibility required patients to be adults 18 years of age or older with a clinical diagnosis of RA by a rheumatologist, fulfilling the American College of Rheumatology/European League Against Rheumatism 2010 revised classification criteria for RA [2]. Patients were excluded if they did not comprehend English, were unable to provide written informed consent, or were members of a vulnerable population (e.g., incarcerated subjects). This led to a total of 76 patients fulfilling the eligibility criteria, who were partitioned into two groups (**Table 1**): for the discovery cohort of this study, 64 patients with available blood samples from at least two outpatient visits 6–12 months apart were included (128 total samples); for the validation cohort, 12 patients whose blood samples were available from only a single outpatient visit were included (12 total samples). Demographic and clinical data, including the numbers of tender and swollen joints, patient and evaluator global assessments, CRP (mg/L), body mass index (BMI, kg/m^2), smoking status, and results for rheumatoid factor (RF, IU/mL) and anti-cyclic citrullinated peptide antibodies (anti-CCP), were collected from the electronic medical records. The patient samples (140 in total) in the study had established disease with mean age 63.54 (range: 32–86), and 69.7% (53 of 76) were female. Disease activity varied from remission to high disease activity, with a DAS28-CRP mean of 3.0 (range: 1.2–7.0). See **Additional file 1** for distribution of DAS28-CRPs corresponding to all study participants.

Table 1. Demographic characteristics of study participants.

	Discovery Cohort$^{\alpha}$		**Validation Cohort$^{\beta}$**
Number of RA patients/samples	64/128		12/12
Sex of RA patients (female/male)	44/20		9/3
	Visit 1	**Visit 2**	-
DAS28-CRP			
Mean ± SD	3.1±1.3	3.0±1.4	2.4±1.3
Range (min–max)	1.5–7.0	1.2–6.6	1.7–5.9
Age (years)			
Mean ± SD	62.7±10.5	63.5±10.6	67.8±10.6
Range (min–max)	32–85	33–86	54–84
BMI			
Mean ± SD	30.6±5.7	31.1±6.2	27.0±4.1
Range (min–max)	22.4–45.3	22.8–47.8	19.0–33.3

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

N/A (n)	6	6	2
Smoking History (n)			
Current (active within 3 months)	7	5	1
Former	31	32	3
Never	25	27	7
N/A	1	0	1
CRP (mg/L)			
Mean ± SD	8.91±16.8	8.0±12.7	11.5±21.7
Range (min–max)	0.29–113	0.7–84	1.0–77.1
RF$^{\gamma}$ (n)			
Positive	36	-	6
Negative	15	-	2
N/A	13	-	4
Anti-CCP$^{\gamma}$ (n)			
Positive	44	-	5
Negative	13	-	1
N/A	7	-	6
Treatment			
Methotrexate use (n, %)	48 (75.0%)	49 (76.6%)	7 (58.3%)
Methotrexate dose (mg/week)			
median	20.0	20.0	22.5
IQR [Q_1, Q_3]	[15.0, 25.0]	[15.0, 25.0]	[17.5,25.0]
Prednisone use (n, %)	29 (45.3%)	28 (43.8%)	4 (33.3%)
Prednisone dose (mg/day)			
median	5.0	5.0	5.0
IQR	[5.0, 7.0]	[5.0, 5.0]	[5.0, 5.0]
TNF*i*-bDMARDs$^{\delta}$ (n, %)	23 (35.9%)	21 (32.8%)	3 (25.0%)
non-TNF*i*-bDMARDs$^{\varepsilon}$ (n, %)	6 (9.4%)	7 (10.9%)	1 (8.3%)
non-methotrexate csDMARDs$^{\lambda}$ (n, %)	20 (31.2%)	27 (42.2%)	1 (8.3%)

$^{\alpha}$Training group. Plasma samples were obtained from patients at two different time-points; $^{\beta}$Test group. Plasma samples were obtained from patients at a single time-point; $^{\gamma}$Reported only for the first visit; $^{\delta}$adalimumab, certolizumab, etanercept, and infliximab; $^{\varepsilon}$abatacept, rituximab, and tocilizumab; $^{\lambda}$azathioprine, hydroxychloroquine, leflunomide, and sulfasalazine; N/A, Not available; RF, rheumatoid factor; Anti-CCP, anti-cyclic citrullinated peptide antibodies; IQR, inter-quartile range; bDMARDs, biologic disease-modifying anti-rheumatic drugs; csDMARDs, conventional synthetic disease-modifying anti-rheumatic drugs; an expanded table with further information on demographic and clinical characteristics is provided in **Additional file 2** and **Additional file 3**.

Metabolomic Profiling

Untargeted metabolomic profiling of plasma samples from both discovery and validation cohorts through ultra-high performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) was performed by Metabolon Inc. (Durham, NC, USA)'s Discovery HD4™ platform. Detailed descriptions of all methods regarding metabolomic profiling are available in **Additional file 4**.

Analysis Workflow

Figure 1 provides a summary of the analytic strategy used on the 128 plasma samples of the discovery cohort to identify associations between metabolites and RA disease activity. The analysis workflow consists of two complementary approaches: Using mixed-effects logistic regression, the first approach identifies metabolites that are differentially abundant between higher and lower disease activity groups, which were determined by DAS28-CRP scores [18-20, 31] (**Fig. 1A**); the second approach uses mixed-effects linear regression to model

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

the relationship between DAS28-CRP and metabolite abundances, allowing the detection of key biochemical features that associate with quantitative disease activity (**Fig. 1B**). To test the predictive accuracy of these selected features when incorporated into a generalized linear model, an additional cohort of twelve plasma metabolomic profiles (from twelve RA patients obtained at single time-points) was collected as an independent validation set.

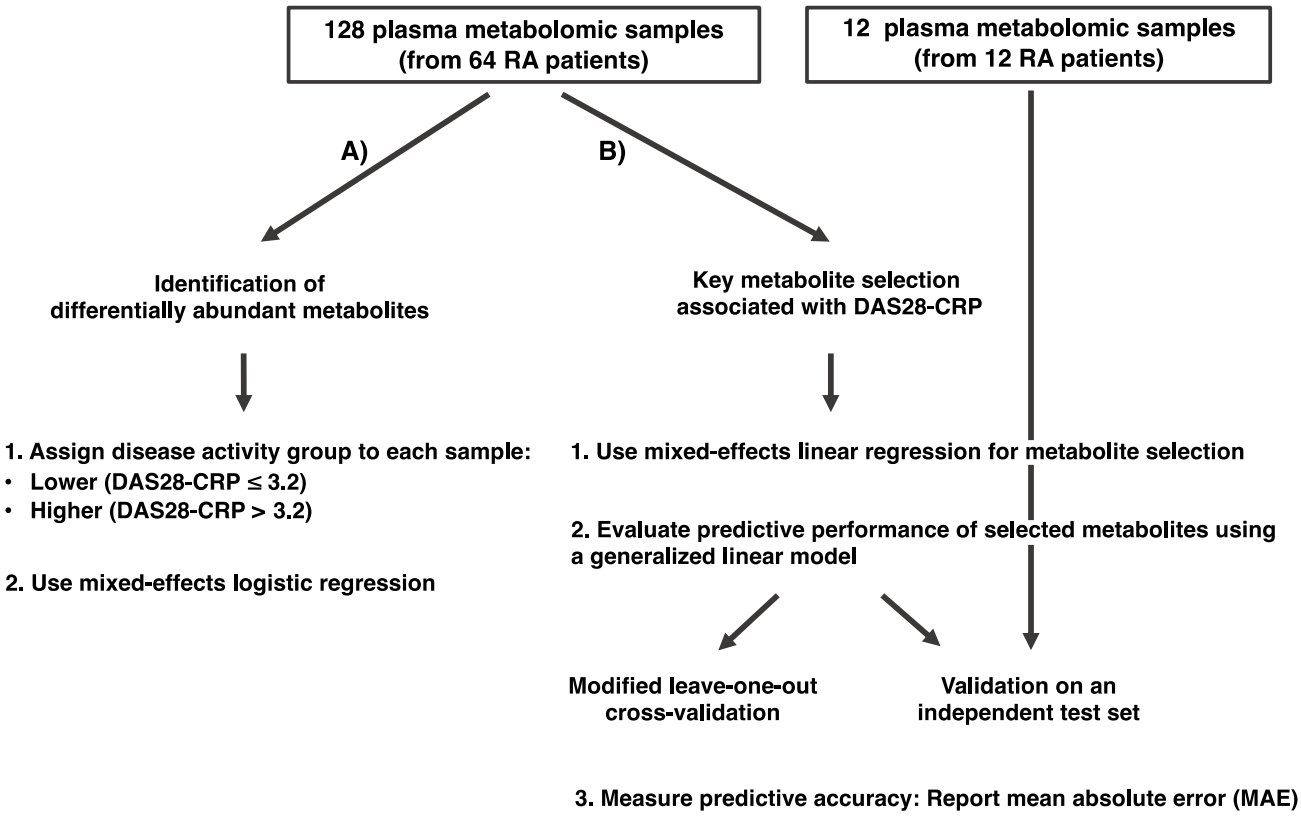

Figure 1. A multi-approach discovery strategy to identify metabolites indicative of RA disease activity. **(A)** Differentially abundant metabolites between higher and lower disease activity groups were identified using a mixed-effects logistic regression model adjusted for patient age and sex, as well as for Patient ID to control for having multiple samples from the same patient. **(B)** A selection scheme to identify metabolites associated with DAS28-CRP. Metabolites were selected with mixed-effects linear regression. To further demonstrate their association with DAS28-CRP, these metabolites were used to construct a generalized linear model for predicting DAS28-CRP. Predictive performance of the model was evaluated on the discovery cohort (using a cross-validation technique) and on a validation cohort.

Pre-processing of Metabolomic Profiling Data

Statistical analyses on untargeted metabolomic data were performed using scaled imputed data provided by Metabolon, Inc. Briefly, the raw data were normalized to account for inter-day variation, which is a result of UPLC-MS/MS runs over multiple days, then the peak intensities were rescaled to set each metabolite's median equal to 1. Missing values were then imputed with the minimum observed value of the metabolite across all samples, finally yielding the scaled imputed data. In addition, metabolites with missing values in over 20% of the entire samples were removed, resulting in 686 metabolites remaining for further analysis. R (v3.6.1), lme4 package (v1.1.21) [32], Python3 (v3.7.5), and sklearn (v0.22.2) were used to perform all data pre-processing and statistical analyses.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Delineation of RA Disease Activity Groups

Following previous reports [18-20, 31], samples from RA patients were divided into two disease activity groups based upon DAS28-CRP: 'lower' (DAS28-CRP ≤ 3.2, n = 76) and 'higher' (DAS28-CRP > 3.2, n = 52). These pre-defined two disease activity groups were used as the nominal response variable in a mixed-effects logistic regression model to identify differentially abundant metabolites between the two groups.

Identification of Differentially Abundant Metabolites While Controlling for Confounding Factors

The following patient characteristics were examined to identify potential confounding factors in the association between plasma metabolites and disease activity (i.e., higher or lower disease activity): age, sex, BMI, smoking history, and treatment use (for methotrexate, prednisone, non-methotrexate csDMARDs, TNF*i*-bDMARDs, and non-TNF*i*-bDMARDS). Based upon the Fisher's exact test, patient age (age ≤ 60, age > 60) and sex (male, female) were observed to have statistically significant associations with the two disease activity groups; the *P*-value for age and sex was $P = 0.01$ (odds ratio [OR] = 2.74, 95% confidence interval [CI] = 1.15–6.73) and $P = 0.02$ (OR = 0.37, 95% CI = 0.14–0.88), respectively. On the other hand, no statistically significant associations were observed between these two disease activity groups and BMI (BMI ≤ 30, BMI > 30; $P = 0.32$), disease duration (duration ≤ 9 years, duration > 9 years; $P = 0.14$), smoking history (smoked at least once, never smoked; $P = 0.36$), or treatment use (user, non-user) for methotrexate ($P = 0.83$), prednisone ($P = 0.58$), TNF*i*-bDMARDs, i.e., adalimumab, certolizumab, etanercept, and infliximab ($P = 0.18$), non-TNF*i*-bDMARDs, i.e., abatacept, rituximab, and tocilizumab ($P = 0.76$), or other non-methotrexate csDMARDs, i.e., azathioprine, hydroxychloroquine, leflunomide, and sulfasalazine ($P = 0.71$). In addition, no significant changes in treatment use were observed between the two visits; *P*-values of the associations between treatment use and time-point based upon McNemar's Chi-squared test for paired nominal data were as follows: methotrexate ($P = 1$), prednisone ($P = 1$), TNF*i*-bDMARDs ($P = 0.75$), non-TNF*i*-bDMARDs ($P = 1$), and non-methotrexate csDMARDs ($P = 0.07$). Therefore, the mixed-effects logistic regression model was adjusted for age and sex as fixed effects, but not for all other aforementioned covariates. In accordance with these results, age and sex have been previously reported to be connected to RA disease activity [33-35]. Herein, patient ID was considered as a random effect in the model to account for intra-subject variance due to having repeated measurements from a single patient. By controlling for patient ID (which are unique to each patient) as a random effect, we are acknowledging the non-independence in our data, that is, sampling that has taken place from within a patient. Leveraging multiple samples from the same patient allows us to compensate for the small number of samples in higher disease activity (DAS28-CRP > 3.2) in each visit (visit 1 and visit 2 having 25 and 27 samples, respectively) by maximizing the degree of freedom for the quantitative disease activity measure, and thereby to boost statistical power. Importantly, no significant difference was observed in DAS28-CRP between visit 1 and visit 2 ($P = 0.98$, Wilcoxon signed-rank test). Metabolites whose corresponding coefficients of the regression model were of *P*-value < 0.05 were considered as differentially abundant, that is, having a statistically significant association with disease activity group.

Selection of Metabolites Associated with DAS28-CRP

Selection of metabolites associated with DAS28-CRP was performed with a mixed-effects linear regression model (DAS28-CRP as the continuous response variable), which controls for fixed effects (scaled metabolite abundances, patients' age and sex) and for random effects (patient ID). Satterthwaite's degrees of freedom

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

method supported by lmerTest (v3.1.1) [36] was applied to test for the statistical significance (*P*-value) of associations between metabolites and DAS28-CRP. *P*-values were retrieved from the corresponding regression coefficients of the predictor variables.

Evaluation of Predictive Performance of DAS28-CRP-associated Metabolites

A generalized linear model (GLM) was used to estimate DAS28-CRP scores using the aforementioned significantly associated metabolites as predictor variables. Predictive performance of the parameterized model was evaluated by two different techniques: First, a modified leave-one-out cross-validation approach was applied to the 128 samples of the training group (discovery cohort). More specifically, in each cross-validation loop, both samples from the same patient were allocated as an internal validation set, while all remaining samples (126 samples from 63 patients) were used to select metabolites significantly associated with DAS28-CRP ($P < 0.05$). These selected biochemical features were then included in a GLM for predicting DAS28-CRP scores of the remaining two samples (of the internal validation group) from their metabolite abundances. The second approach considers testing a GLM, which was composed of the DAS28-CRP-associated metabolites identified from all 128 samples of the training group, on the independent validation group of 12 plasma samples (validation cohort). For both techniques, model performance was reported using mean absolute error (MAE) and standard deviation (SD).

Identification of Metabolites Associated with Treatment Use

A marginal, mixed-effects linear regression model was used to relate metabolite abundance with treatment use. Scaled metabolite abundance, treatment use, and patient ID was set as the response variable, predictor variable (fixed effect), and random effect, respectively. Use of the following treatments was assessed individually: methotrexate, prednisone, non-methotrexate csDMARDs, TNF*i*-bDMARDs, and non-TNF*i*-bDMARDs (names of individual drugs in each treatment group are provided in the footnote of **Table 1**). *P*-values were retrieved from the corresponding regression coefficient of the predictor variable (i.e., use or non-use), and a significance of $P < 0.05$ was reported as statistically significant.

Identification of Differentially Abundant Metabolites Between Two CRP Groups

Metabolites that are significantly associated with disease activity groups and DAS28-CRP scores were further investigated to find those associated with patient groups delineated by CRP levels. First, all samples were divided into two groups as follows: 'high-CRP' (CRP > 3.0 mg/L, n = 52) and 'low-CRP' (CRP ≤ 3.0 mg/L, n = 76). Next, a marginal, mixed-effects linear regression model was used to define the abundance of a metabolite based upon the following fixed effects: CRP group, sex, age, smoking history, and treatment with prednisone, methotrexate, non-methotrexate csDMARDs, TNF*i*-bDMARDs; or non-TNF*i*-bDMARDs. Additionally, patient ID was treated as a random effect. Any covariates whose association with metabolite abundance was statistically significant (i.e., *P*-value of the corresponding regression coefficient < 0.05) were then included in an adjusted mixed model for metabolite abundance. Finally, metabolites were considered as differentially abundant between the two CRP groups if the association between metabolite abundance and CRP group was still found to be significant in the adjusted model ($P < 0.05$).

Results

Differentially Abundant Metabolites between Higher and Lower Disease Activity Groups

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

As shown in our analysis workflow (**Figure 1**), we first sought metabolites that were significantly different in abundance between two major disease activity groups. For this, we divided the 128 metabolomic profiles into two major categories ('higher' vs. 'lower') based upon the reported disease activity of the corresponding patient at the time of sample collection (**Materials and Methods**). Using a mixed-effects logistic regression model (**Materials and Methods**), we identified 33 metabolites as differentially abundant between higher (n = 52) and lower (n = 76) DAS28-CRP groups (**Fig. 2**). Most of these metabolites (31 of 33) were observed to have significantly increased abundances in lower disease activity, whereas the remaining two (glucuronate and hypoxanthine) were found to be significantly increased in higher disease activity. Notably, of the 31 metabolites increased in lower disease activity, seven metabolites (3-hydroxydecanoylcarnitine, dihomo-linoleoylcarnitine (C20:2), eicosenoylcarnitine (C20:1), linoleoylcarnitine (C18:3), linoleoylcarnitine (C18:2), stearoylcarnitine (C18), palmitoylcarnitine (C16)) are a part of acylcarnitine metabolism, and represent a 3.6-fold enrichment in metabolites involved in this particular pathway ($P = 1.9 \times 10^{-3}$, hypergeometric test). It is important to note that the differences seen are relatively small in terms of fold-change, with most of the metabolites varying by 1.1–1.3 fold. Despite these subtle differences within RA patients of varying disease activities, we were still able to obtain statistically significant signal even after considering and controlling for all known potentially confounding factors (which often leads to reduction in statistical power), while adhering to our cut-offs for statistical significance ($P < 0.05$).

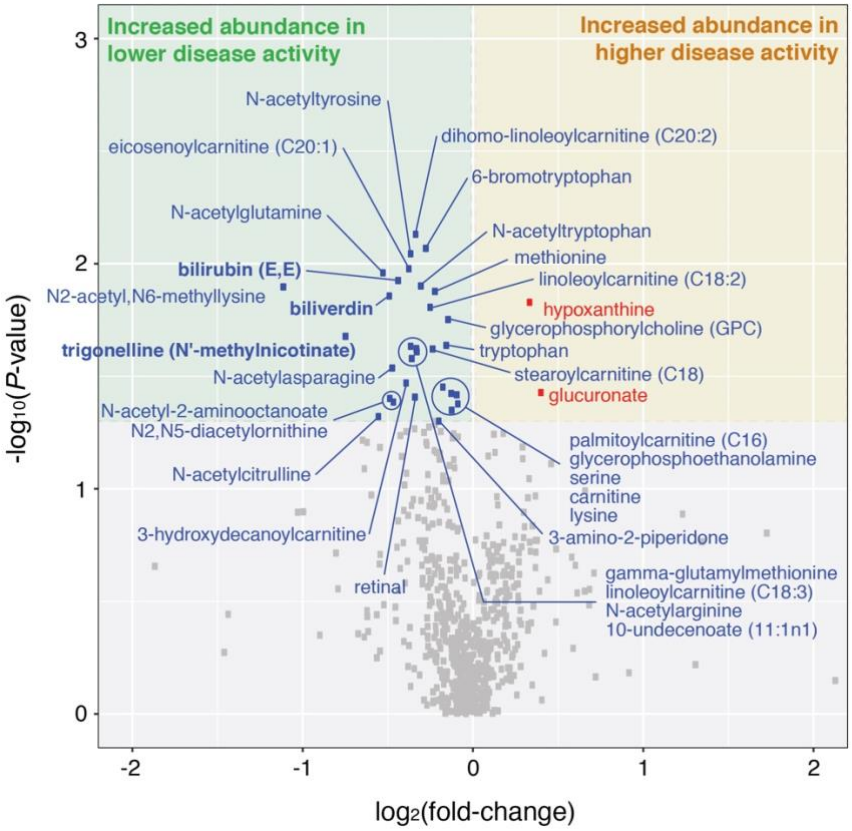

Figure 2. Plasma metabolites differentiating between higher and lower disease activity groups in RA. A total of 2 and 31 metabolites were found to be significantly increased in higher (DAS28-CRP > 3.2, n = 52) and lower (DAS28-CRP ≤ 3.2, n = 76) disease activity groups, respectively. Each point corresponds to a metabolite (686 total). Differentially abundant metabolites were found using a mixed-effects logistic regression model on the discovery cohort (128 samples), for which age and sex were adjusted. Metabolites with *P*-value < 0.05 (based upon the corresponding coefficient of the regression model) were considered as significantly

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

different between groups. *P*-values and fold-changes for all metabolites are listed in **Additional file 5**. Metabolites in bold have been previously described in the literature for their associations with RA.

N2-acetyl,N6-methyllysine ($|\log_2(FC)| = 1.11$, $P = 1.26 \times 10^{-2}$) and trigonelline (N'-methylnicotinate) ($|\log_2(FC)| = 0.74$, $P = 2.09 \times 10^{-2}$), which were both found to have increased abundance in lower disease activity, were the top two metabolites having the largest fold-changes between the two groups. Although the direct relevance of N2-acetyl,N6-methyllysine to RA is currently not well understood, N2-acetyl,N6-methyllysine is part of the lysine metabolism pathway, which has been reported to be associated with RA in the following studies: i) according to Teitsma *et al.*, serum metabolites associated with lysine degradation were observed to have a higher concentration in early RA patients who achieved sustained drug-free remission (after Tocilizumab- or Methotrexate-based treatment) compared to those who never achieved a drug-free status [24]; and ii) Yang *et al.* reported that metabolic products of lysine degradation (carnitine and pipecolic acid) were significantly increased in RA patients than in normal subjects [37].

In regard to trigonelline, which is a product of niacin (vitamin B3) metabolism, this alkaloid has been suggested to have therapeutic potential for diabetes and central nervous system disease [38], and also reported to demonstrate anti-inflammatory properties in mice [39]. In accordance with our results showing decreased abundance of trigonelline in higher disease activity, trigonelline could be of interest in future studies on inflammatory responses in RA.

Biliverdin ($|\log_2(FC)| = 0.48$, $P = 1.38 \times 10^{-2}$) and bilirubin (E,E) ($|\log_2(FC)| = 0.43$, $P = 1.18 \times 10^{-2}$), which are known metabolic products of the heme catabolic pathway, were also observed to have significantly increased abundances in lower disease activity. In particular, biliverdin has been shown to: i) inhibit the activation of pro-inflammatory transcription factors, including NFkB both *in vitro* and *in vivo* [40-44]; ii) inhibit the proliferation of primary T cells stimulated with anti-CD3 and anti-CD28 monoclonal antibodies by inhibiting NFAT/NF-kB activation in a mouse model of heart transplantation [45]; and iii) improve corneal inflammation mediated by heme-oxygenase 2 (HO-2) deficiency in a transgenic mouse model [41]. Moreover, bilirubin, which is derived from the reduction of biliverdin by biliverdin reductase, has been reported as a potential biomarker for RA in line with our findings. For example, Peng *et al.* observed a decreased concentration of serum bilirubin in RA patients compared to healthy controls, as well as in RA patients with worsening disease activity [46]. Additionally, Fischman *et al.* found that total bilirubin levels are inversely related to RA disease activity even after adjusting for multiple confounders (e.g., age, sex, race), and discussed the possibility of bilirubin (a known anti-oxidant) having a physiological anti-inflammatory effect [47]. This point is further elaborated upon by Jangi *et al* [48], who have described in detail the immunosuppressive properties of unconjugated bilirubin in RA and other inflammatory disorders. The full list of differentially abundant metabolites and their associated pathways are shown in **Additional file 5.**

Metabolic Feature Selection Improves DAS28-CRP Prediction Accuracy

Having uncovered metabolites demonstrating altered abundance between two major disease activity groups, we next asked whether quantitative disease activity can be predicted with plasma metabolomes. As untargeted metabolomic profiling can yield a considerable amount of noise and random fluctuations in observed signals [49], it is necessary to first select informative metabolic features that reliably capture relevant aspects of the phenotype of interest [50]. For this, we used mixed-effects linear regression models to select metabolites significantly associated with DAS28-CRP. Afterwards, the abundances of the selected metabolic features were incorporated into a generalized linear model (GLM) to predict DAS28-CRP. For comparison purposes, a GLM was constructed without metabolic feature selection, and thereby taking into consideration all features of a

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

metabolomic profile. Details regarding GLM construction and performance evaluation are provided in **Materials and Methods**.

When applying a modified leave-one-out cross-validation technique to the training group samples (n = 128), we found that the GLM incorporating metabolites that were significantly associated with DAS28-CRP outperformed the model without feature selection (i.e., using all metabolites). As shown in **Figure 3**, the distribution of absolute errors between the observed and predicted DAS28-CRP scores was smaller (with respect to the cumulative area under the error curve) for the GLM with feature selection than that without feature selection. To this point, the prediction MAE (±SD) of the GLM with and without feature selection was 1.51 (±1.89) and 2.02 (±2.52), respectively.

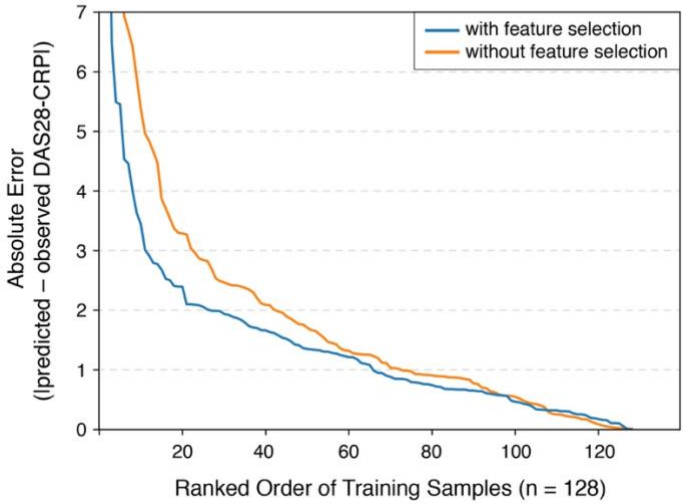

Figure 3. Evaluation of DAS28-CRP predictive performance in cross-validation. A modified leave-one-out cross-validation approach was used on the samples of the training group (128 samples) to test the performance of a generalized linear model (GLM) in predicting DAS28-CRP scores from metabolite abundances. Distributions of absolute errors from models with and without a feature selection scheme were compared to identify the more robust model. The GLM with the feature selection scheme performed better (MAE±SD: 1.51±1.89) than the model without feature selection (MAE±SD: 2.02±2.52).

Having confirmed that feature selection can lead to a more accurate prediction model in cross-validation, we applied the same scheme to all metabolome samples of the discovery cohort to obtain a final set of metabolites associated with DAS28-CRP ($P < 0.05$). After adjusting for potential confounding factors (**Materials and Methods**), this resulted in a collection of 51 plasma metabolites (**Table 2**). These metabolites were used to construct a final GLM, whose predictive accuracy was tested on an independent validation cohort (n = 12) of plasma metabolomic profiles from twelve RA patients (importantly, this additional cohort was not drawn from the same population distribution from which the features were derived). On this previously unseen cohort, the GLM constructed with only the 51 selected metabolites performed considerably better than the model without the feature selection scheme by over two-fold (**Fig. 4A**); the prediction MAE of the GLM with and without feature selection was 0.97 (±0.47) and 2.01 (±2.18), respectively. Likewise, when the actual and predicted DAS28-CRPs were plotted together for both GLMs (**Fig. 4B**), we found that the model with the selection scheme performed more favorably. More specifically, a stronger correlation between the actual and predicted disease activity scores was observed in the model with feature selection (Spearman's $\rho = 0.69$, $P = 1.40 \times 10^{-2}$, 95% CI: [0.18, 0.90]) compared to the model without (Spearman's $\rho = 0.18$, $P = 5.72 \times 10^{-2}$, 95% CI: [-0.44, 0.68]).

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2. Plasma metabolites significantly associated with DAS28-CRP.

Metabolite Name	Super-Pathway[α]	Sub-Pathway[α]	HMDB ID[β]	Regression Coefficient[γ]	*P*-value[δ]
3-hydroxystearate	Lipid	Fatty Acid, Monohydroxy	N/A	0.418	0.002
Phenol sulfate	Amino Acid	Tyrosine Metabolism	HMDB60015	-0.265	0.003
Trimethylamine N-oxide	Lipid	Phospholipid Metabolism	HMDB00925	0.485	0.004
Bilirubin (E,E)	Cofactors and Vitamins	Hemoglobin and Porphyrin Metabolism	N/A	-0.612	0.007
Serine	Amino Acid	Glycine, Serine and Threonine Metabolism	HMDB00187	-1.594	0.010
Dimethylguanidino valeric acid (DMGV)	Amino Acid	Urea cycle; Arginine and Proline Metabolism	N/A	0.325	0.011
N-acetyltryptophan	Amino Acid	Tryptophan Metabolism	HMDB13713	-0.918	0.012
Glycoursodeoxycholate	Lipid	Secondary Bile Acid Metabolism	HMDB00708	0.051	0.012
N-acetylneuraminate	Carbohydrate	Aminosugar Metabolism	HMDB00230	0.470	0.013
Dihomo-linoleoylcarnitine (C20:2)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Polyunsaturated)	N/A	-0.745	0.013
N-acetyltyrosine	Amino Acid	Tyrosine Metabolism	HMDB00866	-0.713	0.014
Branched chain 14:0 dicarboxylic acid	Lipid	Fatty Acid, Dicarboxylate	N/A	-0.201	0.014
1-carboxyethylvaline	Amino Acid	Leucine, Isoleucine and Valine Metabolism	N/A	0.408	0.015
(14 or 15)-methylpalmitate (a17:0 or i17:0)	Lipid	Fatty Acid, Branched	N/A	0.227	0.017
Isoursodeoxycholate	Lipid	Secondary Bile Acid Metabolism	HMDB00686	0.059	0.018
Glucuronate	Carbohydrate	Aminosugar Metabolism	HMDB00127	0.396	0.019
Glucose	Carbohydrate	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism	HMDB00122	1.107	0.019
Linoleoylcarnitine (C18:3)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Polyunsaturated)	N/A	-0.534	0.020
1-methylhistidine	Amino Acid	Histidine Metabolism	HMDB00001	0.580	0.020
Trigonelline (N'-methylnicotinate)	Cofactors and Vitamins	Nicotinate and Nicotinamide Metabolism	HMDB00875	-0.227	0.020
Palmitoyl ethanolamide	Lipid	Endocannabinoid	HMDB02100	0.067	0.020
Hypoxanthine	Nucleotide	Purine Metabolism, (Hypo)Xanthine/Inosine containing	HMDB00157	0.482	0.022
Biliverdin	Cofactors and Vitamins	Hemoglobin and Porphyrin Metabolism	HMDB01008	-0.436	0.022
Linoleoylcarnitine (C18:2)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Polyunsaturated)	HMDB06469	-0.814	0.023
3-methylhistidine	Amino Acid	Histidine Metabolism	HMDB00479	0.140	0.025
N-acetylarginine	Amino Acid	Urea cycle; Arginine and Proline Metabolism	HMDB04620	-0.755	0.026
4-guanidinobutanoate	Amino Acid	Guanidino and Acetamido Metabolism	HMDB03464	0.347	0.026
1-carboxyethylisoleucine	Amino Acid	Leucine, Isoleucine and Valine Metabolism	N/A	0.307	0.026
Cysteinylglycine disulfide	Amino Acid	Glutathione Metabolism	HMDB00709	1.562	0.027
Guanidinoacetate	Amino Acid	Creatine Metabolism	HMDB00128	-1.125	0.027
N2-acetyl,N6-Methyllysine	Amino Acid	Lysine Metabolism	N/A	-0.213	0.028
Lysine	Amino Acid	Lysine Metabolism	HMDB00182	-1.395	0.031
1,6-anhydroglucose	Xenobiotics	Food Component/Plant	HMDB00640	0.097	0.032
Pyrraline	Xenobiotics	Food Component/Plant	HMDB33143	0.190	0.032

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Mannose	Carbohydrate	Fructose, Mannose and Galactose Metabolism	HMDB00169	0.633	0.032
Ectoine	Xenobiotics	Chemical	N/A	0.123	0.036
6-bromotryptophan	Amino Acid	Tryptophan Metabolism	N/A	-0.758	0.037
1-linoleoyl-GPA (18:2)	Lipid	Lysophospholipid	HMDB07856	-0.371	0.039
Eicosenoylcarnitine (C20:1)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Monounsaturated)	N/A	-0.557	0.039
Erucate (22:1n9)	Lipid	Long Chain Monounsaturated Fatty Acid	HMDB02068	0.346	0.040
Bilirubin	Cofactors and Vitamins	Hemoglobin and Porphyrin Metabolism	HMDB00054	-0.432	0.042
Stearoyl ethanolamide	Lipid	Endocannabinoid	HMDB13078	0.070	0.043
3-phenylpropionate (hydrocinnamate)	Xenobiotics	Benzoate Metabolism	HMDB00764	-0.178	0.043
beta-hydroxyisovalerate	Amino Acid	Leucine, Isoleucine and Valine Metabolism	HMDB00754	0.723	0.045
Myo-inositol	Lipid	Inositol Metabolism	HMDB00211	0.944	0.045
Gulonate	Cofactors and Vitamins	Ascorbate and Aldarate Metabolism	HMDB03290	0.575	0.047
Gluconate	Xenobiotics	Food Component/Plant	HMDB00625	0.539	0.047
Tryptophan	Amino Acid	Tryptophan Metabolism	HMDB00929	-1.139	0.048
1-carboxyethylleucine	Amino Acid	Leucine, Isoleucine and Valine Metabolism	N/A	0.350	0.048
alpha-ketobutyrate	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism	HMDB00005	0.268	0.049
Lanthionine	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism	N/A	-0.229	0.049

[α]Super-pathways and sub-pathways were defined by Metabolon's Discovery HD4™ platform; [β]Metabolite IDs provided by the Human Metabolome Database (HMDB); [γ]Coefficients of the predictor variables (metabolites) in the mixed-effects linear regression model from discovery cohort (n = 128). Sign and magnitude of the coefficient indicate direction and strength of the correlation (between the metabolite and DAS28-CRP), respectively; [δ]*P*-values were retrieved from the corresponding regression coefficients; N/A, not available.

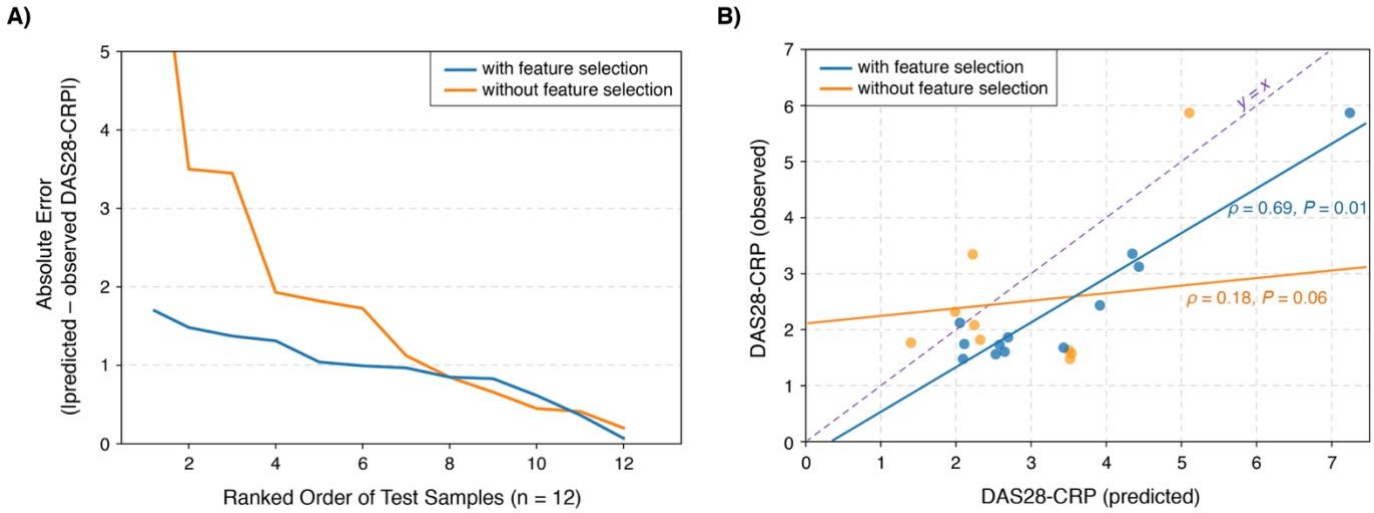

Figure 4. GLM with feature selection provides improved DAS28-CRP prediction accuracy in an independent validation group (12 samples). (A) Performance of GLMs in predicting quantitative disease activity were evaluated on samples of an independent validation group. Distributions of absolute errors from models with and without a feature selection scheme were compared to identify the more robust model. **(B)** Selection of metabolic features prior to model training resulted in higher predictive performance, as evidenced by the stronger correlation between observed and predicted DAS28-CRPs. Three samples predicted to have negative DAS28-CRP values are omitted from the scatter-plot. Dashed violet line indicates 'y = x', i.e., an exact match between

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

the observed and predicted values. 95% confidence interval for ρ with feature selection: [0.18, 0.90]; without feature selection: [-0.44, 0.68].

Commonly Identified Metabolites from Two Different Analytic Approaches

To summarize the findings above, we found that, from the 686 total detectable metabolites in a metabolomic profile, 33 (4.8%) were differentially abundant between higher and lower disease activity; and 51 (7.4%) were significantly associated with DAS28-CRP (**Fig. 5**). These separate findings amounted to a total of 67 unique metabolites, among which 40 were found to have no association with the use of prednisone, methotrexate, other non-methotrexate csDMARDs, TNF*i*-bDMARDs, or non-TNF*i*-bDMARDs (**Materials and Methods**). Notably, eight metabolites (6-bromotryptophan, bilirubin (E,E), biliverdin, glucuronate, N-acetyltryptophan, N-acetyltyrosine, serine, and trigonelline) were not only consistently detected across both analytic approaches, but also found to have no association with any treatment use; these results strongly suggest key metabolic pathways and modules potentially contributing to, or serving as indicators of, RA pathogenesis independent of confounding treatment effects. Consistent with this idea, additional studies into the metabolites found in this study (the majority of which have yet to be linked to RA) may be able to provide new insight into the perturbed physiological metabolic processes—which are then in turn reflected in blood—underlying disease progression in RA.

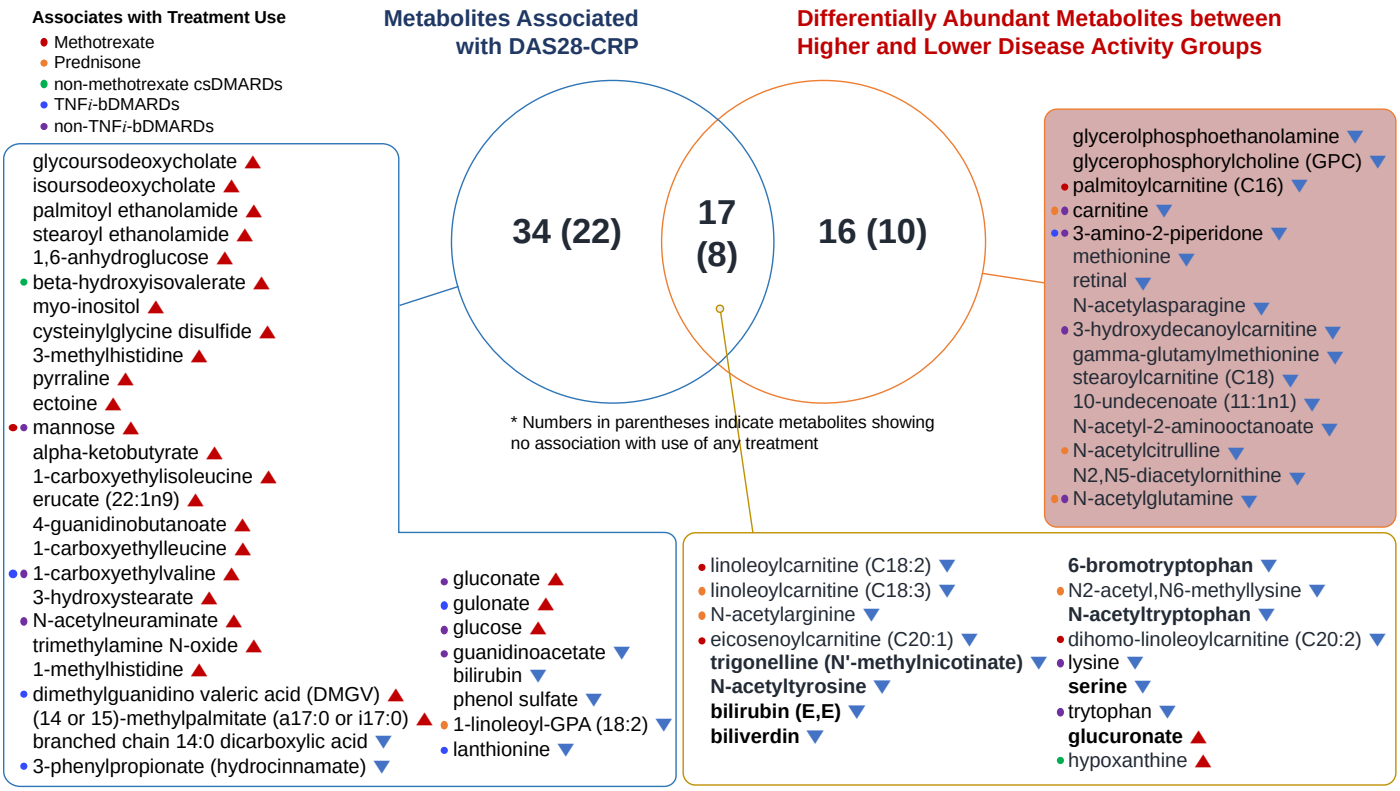

Figure 5. Venn diagram of all plasma metabolites identified through the multi-approach discovery strategy. A total of 67 unique metabolites were identified, among which 40 were found to have no association with the use of treatment. Notably, eight metabolites (6-bromotryptophan, bilirubin (E,E), biliverdin, glucuronate, N-acetyltryptophan, N-acetyltyrosine, serine, and trigonelline) in bold were not only consistently detected across both analytic approaches, but also found to have no association with any treatment use. Colored circles indicate metabolites whose abundances associate with treatment use. Metabolites with red triangles were found to have increasing abundances with worsening disease activity, whereas metabolites with blue triangles were found to have decreasing abundances with worsening disease activity.

Metabolites Associated with CRP Patient Groups

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Elevated levels of C-reactive protein (CRP) in the blood is well known to often indicate increased inflammatory conditions, which may be caused by a wide variety of acute (e.g., infections) and chronic disorders (e.g., rheumatoid arthritis, inflammatory bowel disease). In RA patients, CRP levels have been observed to increase after acute mental stress tasks [51], and also to be linked to risk of cardiovascular disease [52]. Furthermore, several serum metabolites were found to reflect inflammatory activity in patients with early arthritis [53].

We further investigated the aforementioned 67 plasma metabolites to see whether any were differentially abundant between two CRP patient groups, i.e., 'high-CRP' (CRP > 3.0 mg/L, n = 52) and 'low-CRP' (CRP ≤ 3.0 mg/L, n = 76) (**Materials and Methods**). While controlling for potential confounding variables, we identified eight total metabolites that were significantly associated with CRP patient group. More specifically, the abundances of mannose, beta-hydroxyisovalerate, (14 or 15)-methylpalmitate (a17:0 or i17:0), erucate (22:1n9), 10-undecenoate (11:1n1), N-acetylcitrulline were higher in high-CRP, while those of serine and linoleoylcarnitine (C18:3) were lower in high-CRP (**Fig. 6**). Application of these plasma metabolites, which were found to be connected to both RA disease activity and circulating CRP levels, may lead to the development of new clinical laboratory tests to further enable precision medicine for RA patients.

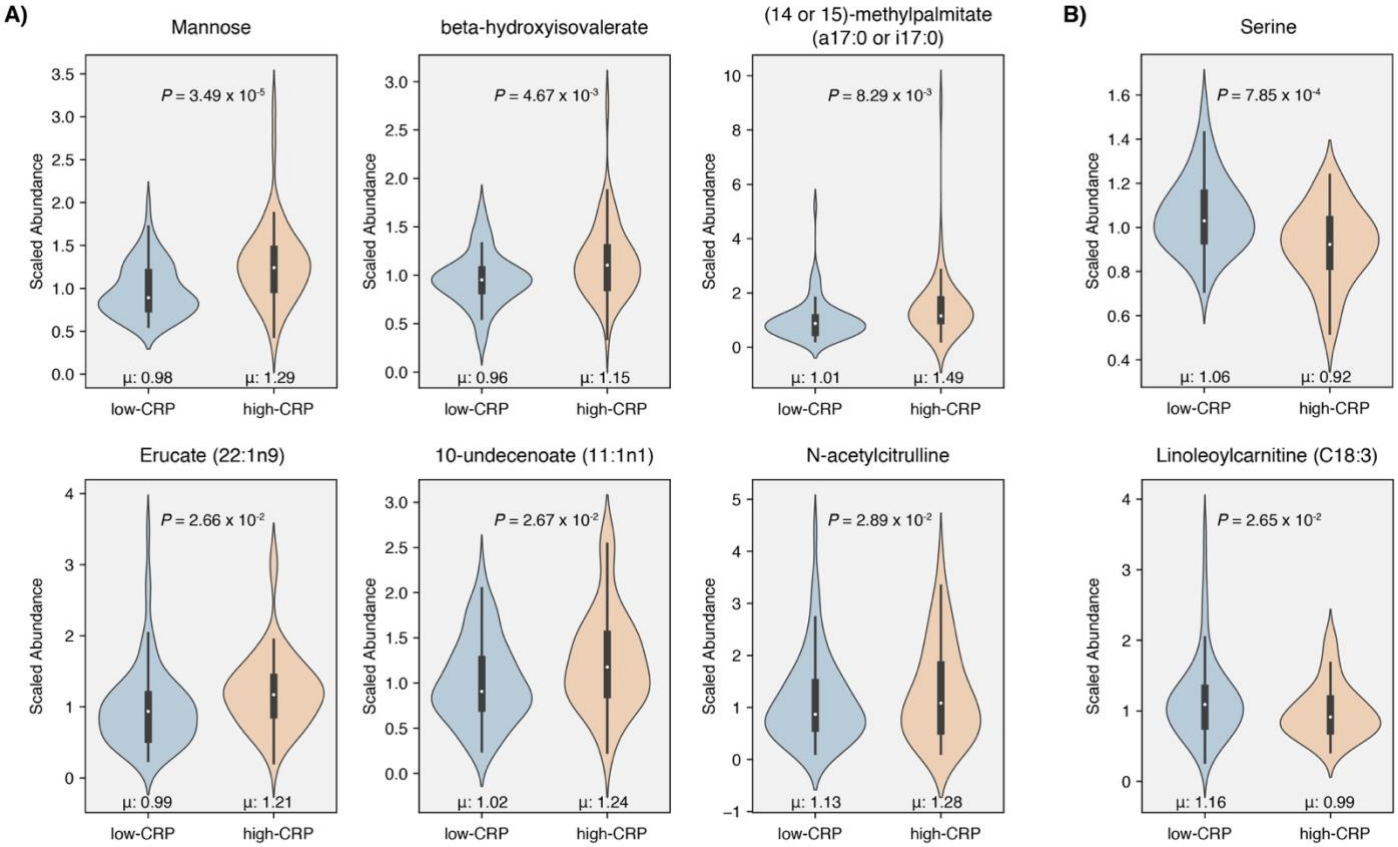

Figure 6. Metabolites differentially abundant between two CRP patient groups. Among the 67 total metabolites identified through our multi-approach analysis on the discovery cohort (n = 128), eight metabolites were identified to have significant associations with CRP group while controlling for confounding variables (regression coefficient for CRP, $P < 0.05$) **(A)** Metabolites with higher abundances in the high-CRP group: mannose, beta-hydroxyisovalerate, (14 or 15)-methylpalmitate (a17:0 or i17:0), erucate (22:1n9), 10-undecenoate (11:1n1), and N-acetylcitrulline. **(B)** Metabolites with higher abundances in the low-CRP group: serine and linoleoylcarnitine (C18:3).

Plasma metabolites associate with clinical improvement in RA

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Based upon the European League Against Rheumatism (EULAR) response criteria for DAS28-CRP [54], we found that sixteen of the 64 patients in the discovery cohort showed moderate or good improvement in disease activity from visit 1 to visit 2, while the remaining 48 patients did not show clinical improvement at the time of their second visit. This discovery provided an entry point for the following analysis: For each of these two patient groups, i.e., 'Improved' (n = 16) and 'Non-improved' (n = 48) patients, we aimed to identify metabolites whose abundances significantly changed from visit 1 to visit 2, while controlling for the same confounding factors (mixed-effects regression model, $P < 0.05$). As a result, we identified eleven metabolites whose abundances significantly changed in the Improved patient group (**Additional file 6)**, while nineteen metabolites showed significant changes in the Non-improved patient group (**Additional file 7)**. The following three metabolites, which were discovered in our previous analyses on the 128 plasma metabolome samples of the discovery cohort, were detected once again: erucate (22:1n9), a metabolite identified to be associated with both DAS28-CRP and CRP patient group, was identified to be significantly different between visit 1 and visit 2 in patients who did not show clinical improvement (Non-improved); 3-amino-2-piperidone, a metabolite identified to be differentially abundant between higher and lower disease activity in our study, was identified to be significantly different between visit 1 and visit 2 in patients who showed clinical improvement (Improved); and gamma-glutamylmethionine, a metabolite identified to be differentially abundant between higher and lower disease activity, was identified to be significantly different between visit 1 and visit 2 in the Non-improved group. These results allow us to expand our future direction to investigate metabolites associated with clinical improvement in patients with RA.

Discussion

Dysfunctions in cellular and tissue metabolism are involved in a broad range of autoimmune disorders [55-58], including RA [59-61]. These metabolic implications highlight the importance of investigating which biochemical functions and metabolic states are altered during the onset and progression of disease. To this end, metabolomics platforms (and the accompanying wealth of data) can present unique opportunities for discovering novel disease 'molecular signatures' [62], which can be interpreted through the lens of annotated biochemical relationships. Moreover, high-throughput profiling can enable the identification of circulating pro-inflammatory (disease-triggering) and anti-inflammatory (disease-protective) metabolites in RA, as elaborated upon by Coras *et al.* [63]; such discoveries may facilitate the design of either dietary or gut microbiome-based intervention strategies to improve wellness or alter the course of disease for RA patients. In this study, by performing a global metabolomic profiling analysis on 128 plasma samples obtained from 64 patients with RA, we identified biochemical signatures associated with, and predictive of, disease activity. Mainly, through a combination of statistical approaches for metabolic signature discovery, we identified several metabolites that: i) differ significantly between lower and higher disease activity groups; and ii) significantly associate with DAS28-CRP. Of note, our study is the first to leverage biochemical features from a plasma metabolomic profile to predict quantitative disease activity.

Interestingly, we identified eight metabolites (6-bromotryptophan, bilirubin (E,E), biliverdin, glucuronate, N-acetyltryptophan, N-acetyltyrosine, serine, and trigonelline) that were commonly found across different statistical approaches, possibly capturing representative metabolite signals of RA progression. We discussed above the reported roles of bilirubin and biliverdin in RA. Moreover, these two metabolites were previously reported for their cytoprotective and anti-inflammatory effects [64-68], and even suggested as an "RA protective factor" by Fischman *et al.* [47]. Interestingly, high concentrations of bilirubin and biliverdin were reported in other inflammatory disorders, such as atherosclerotic diseases [69] and autoimmune encephalitis

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

[67]. In regards to the other six metabolites, clear and definitive connections with RA have not yet been established. However, if our results on bilirubin and biliverdin were to serve as benchmarks for reliably identifying plasma metabolites important to RA disease activity, then these remaining metabolites may be deemed as leading candidates for future investigations.

Glucuronate was found to show elevated abundance in higher disease activity than in lower disease activity. This glucose-derivative is involved in the detoxification of xenobiotics via glucuronidation in the mammalian liver. Interestingly, this process can be reversed by gut bacteria harboring β-D-glucuronidases [70], and thereby releasing (potentially toxic) exogenous compounds into the gut lumen and subsequently into circulation [71, 72]. In that respect, examining a possible role involving dysbiosis in the gut microbiota—combined with metabolomic approaches to infer relationships between gut microbes and blood metabolites in RA, as shown by Chen *et al.* [73]—may help elucidate a microbial-based mechanism explaining the observed alterations in plasma glucuronate.

Serine was seen to decrease with worsening disease activity. In line with our results, albeit in an RA mouse model with collagen-induced arthritis (CIA), plasma levels of serine and other free amino acids were found to have significantly decreased in the CIA group compared to control mice [74]. In another study wherein synovial fluid of RA patients were examined for citrullinated proteins (which is widely known to result in a rise in anti-citrullinated protein antibodies in RA), Tilvawala *et al.* found increased citrullination in a wide array of serine protease inhibitors (Serpins) and serine proteases [75]; in the same study, the investigators demonstrated *in vitro* that citrullinating serine protease inhibitors nearly abolishes their inhibitory activity towards their target proteases. Although we have yet to uncover whether a decrease in plasma serine levels (with worsening disease activity) is linked to citrullination of serine proteases in synovial fluid, we speculate that changes in serine may reflect dysregulated protein degradation during systemic inflammatory activity and joint destruction in RA.

We note a few limitations of this study: First, we acknowledge that our study includes a relatively small number of samples within each disease activity group of the discovery cohort and the validation cohort. Nevertheless, we were able to detect statistically significant metabolites in all analytical strategies, demonstrating that our data provides reasonably sufficient statistical power. Certainly, a much larger cohort would have been ideal; however, this is a small pilot study on our stored plasma samples, and obtaining an additional cohort is outside of its scope. Encouragingly, despite the low sample size of the validation cohort, we were able to successfully show that feature selection is a necessary step in the model-training process, and we expect this finding to translate well to larger cohorts in our future studies. Nevertheless, in order to more meticulously examine the role of blood metabolites in RA, future investigations will warrant a higher number of samples and more detailed subject characteristics. Second, to define RA disease activity, we solely used the DAS28-CRP scoring system, which is dependent upon acute-phase responses that may not accurately reflect patients who have an inflammation-free state [24]. Our future plans include performing our analytical pipeline with other RA disease activity metrics (e.g., clinical disease activity index (CDAI), simple disease activity index (SDAI)) to test the robustness of our findings. Third, all of our multivariate analyses followed adjustment for patient age and sex only. Other potential confounders that may affect the concentration of blood metabolites, such as diet, exercise habits, lifestyle factors, time of the day of sample collection, and gut microbiome, were not considered as predictor variables in our analyses. Fourth, comorbidity can certainly be a significant confounding factor. Alternatively, comorbidities could theoretically be contributors or mediators to inflammatory disease activity in patients with RA. At this stage, it would have been premature to adjust statistically for the effects of particular comorbidities or for the presence of multi-morbidity (i.e., multiple chronic conditions) before carefully investigating for their potential interaction with plasma metabolites and

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

RA disease activity. This may lead to 'over-adjustment' and falsely concluding that certain metabolites are not significant when they may, in fact, be very important. Future studies will be necessary to explore potential interactions between comorbidities, RA disease activity, and plasma metabolite levels in RA. Last, despite the similarities in our findings with previous investigations (as noted above), many of our results are reported for the first time and remain to be validated by others. Possible causes of discrepancies with the work of others include the comparatively small number of samples in this pilot study; technical and biological sources of random noise; the uniqueness of our recruited patient cohort; variabilities in detection protocols and instrument sensitivity; and use of alternate statistical techniques and potential over-fitting. Future efforts, by us and others, are likely to elucidate truly robust signal and collectively strengthen the confidence in our novel findings.

Despite the aforementioned limitations, our study establishes the far-reaching utility of using cutting-edge technological and analytical approaches for plasma metabolomic profiling and justifies analogous investigations at larger scales. The identified metabolites could be a reflection of the perturbed metabolic processes concurrent with worsening disease activity, and our findings will inspire future studies into how inflammation and pain in RA are coupled to physiological metabolism. Moreover, our identified sets of signature metabolites offer a promising glimpse into biomolecular marker panels for diagnosing disease activity of RA patients solely through blood (thereby complementing current diagnostic approaches), with the overall aim to make such assessments faster, cheaper, and less invasive. In turn, studies such as ours are expected to contribute towards fully realizing the potential of virtual and digital healthcare by foregoing the need for patients to physically arrive at the clinic to meet their primary care provider in person.

As the gut microbiome has been recognized to be implicated in RA [73, 76-78]—possibly through complex mechanisms underlying microbe-microbe and host-microbe biochemical cross-talk [79]—integrated profiling of both stool metagenome and blood metabolome would provide an in-depth, comprehensive view of functional dysbiosis during RA onset and progression. Interestingly, a recent study showed that blood metabolites can be predictive of gut microbiome alpha-diversity [80]. Such investigations into integrating across multiple data types of the same phenotype can help to amplify the primary biological signal of interest relative to noise, as well as provide actionable insights. In conclusion, the results reported herein are poised to eventually improve disease management and outcomes of patients with RA and other rheumatic diseases, as well as to provide novel means of monitoring health and wellness [81].

List of Abbreviations

RA: Rheumatoid arthritis; DMARDs: Disease-modifying anti-rheumatic drugs; bDMARDs: Biologic disease-modifying anti-rheumatic drugs; csDMARDs: Conventional synthetic disease-modifying anti-rheumatic drugs; UPLC-MS/MS: Ultra-high performance liquid chromatography-tandem mass spectrometry; DAS28-CRP: Disease Activity Score-28 using C-reactive protein; BMI: Body mass index; GLM: Generalized linear model; MAE: Mean absolute error; SD: Standard deviation; ESR: Erythrocyte sedimentation rate; CRP: C-reactive protein; Anti-CCP: anti-cyclic citrullinated peptide antibodies; HMDB: Human metabolome database; CIA: Collagen-induced arthritis; CDAI: Clinical disease activity index; SDAI: Simple disease activity index.

Supplementary Information

Additional file 1: Supplementary Figure 1, Histogram of DAS28-CRPs corresponding to the 128 total samples of the discovery cohort.

Additional file 2: Supplementary Table 1, Subject characteristics of 64 patients (128 samples) from the discovery cohort.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Additional file 3: Supplementary Table 2, Subject characteristics of 12 patients (12 samples) from the validation cohort.
Additional file 4: Supplementary Information on detailed methods regarding metabolomic profiling (Metabolon).
Additional file 5: Supplementary Table 3, Differentially abundant metabolites between lower (DAS28-CRP ≤ 3.2) and higher (DAS28-CRP > 3.2) disease activity groups.
Additional file 6: Supplementary Table 4, Metabolites displaying significant changes in abundances in patients with clinical improvement.
Additional file 7: Supplementary Table 5, Metabolites displaying significant changes in abundances in patients without clinical improvement.

Declarations

Ethics approval and consent to participate

This study was approved by the Mayo Clinic Institutional Review Board (no. 14-000616 and no. 14-000680) in accordance with the Declaration of Helsinki. All methods and procedures were performed in accordance with Mayo Clinic Institutional Review Board guidelines and regulations.

Consent for publication

All patients provided written informed consent.

Availability of data and materials

All raw metabolomic datasets, as well as source codes used to reproduce all results in this study, are available at https://github.com/jaeyunsung/RA_plasma_metabolomics_2020.

Competing interests

The authors declare no competing interests.

Funding

This work was supported in part by the Mayo Clinic Center for Individualized Medicine (to B.H., V.K.G., and J.S.), and Mark E. and Mary A. Davis to Mayo Clinic Center for Individualized Medicine (J.M.D. and J.S.).

Authors' contributions

B.H., J.M.D., and J.S. conceived the problem and designed all analytical methodologies. B.H. performed all computational analyses. All authors analyzed the data. B.H., J.M.D., and J.S. wrote the manuscript, with editorial contributions from other authors. J.M.D. is the principal investigator of the Mayo Clinic Rheumatology Biobank, from which blood samples were collected from patients with rheumatoid arthritis. All authors reviewed and approved the final manuscript.

Acknowledgements

First and foremost, we thank our dear patients who volunteered for this study. We also thank the Mayo Clinic Division of Rheumatology study coordinators (Jennifer Sletten and Kathleen McCarthy-Fruin) for their help in making this work possible.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Guo Q, Wang Y, Xu D, Nossent J, Pavlos NJ, Xu J: **Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies**. *Bone research* 2018, **6**(1):1-14.
2. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham III CO, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD: **2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative**. *Arthritis & rheumatism* 2010, **62**(9):2569-2581.
3. Smolen JS, Aletaha D, Barton A, Burmester GR, Emery P, Firestein GS, Kavanaugh A, McInnes IB, Solomon DH, Strand V *et al*: **Rheumatoid arthritis**. *Nature Reviews Disease Primers* 2018, **4**(1):18001.
4. Firestein GS: **Evolving concepts of rheumatoid arthritis**. *Nature* 2003, **423**(6937):356-361.
5. Aletaha D, Smolen JS: **Diagnosis and management of rheumatoid arthritis: a review**. *Jama* 2018, **320**(13):1360-1372.
6. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, Silman AJ: **Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins**. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* 2000, **43**(1):30-37.
7. Crowson CS, Matteson EL, Myasoedova E, Michet CJ, Ernste FC, Warrington KJ, Davis III JM, Hunder GG, Therneau TM, Gabriel SE: **The lifetime risk of adult-onset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases**. *Arthritis & Rheumatism* 2011, **63**(3):633-639.
8. Ngo ST, Steyn FJ, McCombe PA: **Gender differences in autoimmune disease**. *Frontiers in neuroendocrinology* 2014, **35**(3):347-369.
9. Sugiyama D, Nishimura K, Tamaki K, Tsuji G, Nakazawa T, Morinobu A, Kumagai S: **Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies**. *Annals of the rheumatic diseases* 2010, **69**(01):70-81.
10. Khanna S, Jaiswal KS, Gupta B: **Managing rheumatoid arthritis with dietary interventions**. *Frontiers in nutrition* 2017, **4**:52.
11. Julià A, Alonso A, Marsal S: **Metabolomics in rheumatic diseases**. *International Journal of Clinical Rheumatology* 2014, **9**(4):353.
12. Aletaha D, Smolen J: **The rheumatoid arthritis patient in the clinic: comparing more than 1300 consecutive DMARD courses**. *Rheumatology* 2002, **41**(12):1367-1374.
13. Yap H-Y, Tee SZ-Y, Wong MM-T, Chow S-K, Peh S-C, Teow S-Y: **Pathogenic role of immune cells in rheumatoid arthritis: implications in clinical treatment and biomarker development**. *Cells* 2018, **7**(10):161.
14. Haringman JJ, Gerlag DM, Zwinderman AH, Smeets TJ, Kraan MC, Baeten D, McInnes IB, Bresnihan B, Tak P: **Synovial tissue macrophages: a sensitive biomarker for response to treatment in patients with rheumatoid arthritis**. *Annals of the rheumatic diseases* 2005, **64**(6):834-838.
15. Chung S-J, Kwon Y-J, Park M-C, Park Y-B, Lee S-K: **The correlation between increased serum concentrations of interleukin-6 family cytokines and disease activity in rheumatoid arthritis patients**. *Yonsei medical journal* 2011, **52**(1):113-120.
16. Arts EE, Fransen J, Den Broeder AA, van Riel PL, Popa CD: **Low disease activity (DAS28≤ 3.2) reduces the risk of first cardiovascular event in rheumatoid arthritis: a time-dependent Cox regression analysis in a large cohort study**. *Annals of the Rheumatic Diseases* 2017, **76**(10):1693-1699.
17. Van Gestel A, Prevoo M, Van't Hof M, Van Rijswijk M, Van de Putte L, Van Riel P: **Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis: comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism criteria**. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* 1996, **39**(1):34-40.
18. Prevoo M, Van'T Hof MA, Kuper H, Van Leeuwen M, Van De Putte L, Van Riel P: **Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective**

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

longitudinal study of patients with rheumatoid arthritis. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* 1995, **38**(1):44-48.

19. Fransen J, Van Riel P: **The Disease Activity Score and the EULAR response criteria**. *Clinical and experimental rheumatology* 2005, **23**(5):S93.
20. Anderson J, Caplan L, Yazdany J, Robbins ML, Neogi T, Michaud K, Saag KG, O'dell JR, Kazi S: **Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice**. *Arthritis care & research* 2012, **64**(5):640-647.
21. Li J, Che N, Xu L, Zhang Q, Wang Q, Tan W, Zhang M: **LC-MS-based serum metabolomics reveals a distinctive signature in patients with rheumatoid arthritis**. *Clinical rheumatology* 2018, **37**(6):1493-1502.
22. Patti GJ, Yanes O, Siuzdak G: **Metabolomics: the apogee of the omics trilogy**. *Nature reviews Molecular cell biology* 2012, **13**(4):263-269.
23. Clish CB: **Metabolomics: an emerging but powerful tool for precision medicine**. *Molecular Case Studies* 2015, **1**(1):a000588.
24. Teitsma XM, Yang W, Jacobs JW, Pethö-Schramm A, Borm ME, Harms AC, Hankemeier T, van Laar JM, Bijlsma JW, Lafeber FP: **Baseline metabolic profiles of early rheumatoid arthritis patients achieving sustained drug-free remission after initiating treat-to-target tocilizumab, methotrexate, or the combination: insights from systems biology**. *Arthritis research & therapy* 2018, **20**(1):230.
25. Sasaki C, Hiraishi T, Oku T, Okuma K, Suzumura K, Hashimoto M, Ito H, Aramori I, Hirayama Y: **Metabolomic approach to the exploration of biomarkers associated with disease activity in rheumatoid arthritis**. *PloS one* 2019, **14**(7).
26. Hambardzumyan K, Bolce RJ, Wallman JK, Van Vollenhoven RF, Saevarsdottir S: **Serum biomarkers for prediction of response to methotrexate monotherapy in early rheumatoid arthritis: Results from the SWEFOT trial**. *The Journal of rheumatology* 2019, **46**(6):555-563.
27. Tasaki S, Suzuki K, Kassai Y, Takeshita M, Murota A, Kondo Y, Ando T, Nakayama Y, Okuzono Y, Takiguchi M: **Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission**. *Nature communications* 2018, **9**(1):1-12.
28. Kapoor SR, Filer A, Fitzpatrick MA, Fisher BA, Taylor PC, Buckley CD, McInnes IB, Raza K, Young SP: **Metabolic profiling predicts response to anti–tumor necrosis factor α therapy in patients with rheumatoid arthritis**. *Arthritis & Rheumatism* 2013, **65**(6):1448-1456.
29. Cuppen BV, Fu J, Van Wietmarschen HA, Harms AC, Koval S, Marijnissen AC, Peeters JJ, Bijlsma JW, Tekstra J, Van Laar JM: **Exploring the inflammatory metabolomic profile to predict response to TNF-α inhibitors in rheumatoid arthritis**. *PloS one* 2016, **11**(9):e0163087.
30. Liebal UW, Phan AN, Sudhakar M, Raman K, Blank LM: **Machine Learning Applications for Mass Spectrometry-Based Metabolomics**. *Metabolites* 2020, **10**(6):243.
31. Welsing PM, van Riel PL: **The Nijmegen inception cohort of early rheumatoid arthritis**. *The Journal of Rheumatology Supplement* 2004, **69**:14-21.
32. Bates D, Mächler M, Bolker B, Walker S: **Fitting linear mixed-effects models using lme4**. *arXiv preprint arXiv:14065823* 2014.
33. Siemons L, ten Klooster PM, Vonkeman HE, van Riel PL, Glas CA, van de Laar MA: **How age and sex affect the erythrocyte sedimentation rate and C-reactive protein in early rheumatoid arthritis**. *BMC musculoskeletal disorders* 2014, **15**(1):368.
34. Radovits B, Fransen J, Van Riel P, Laan R: **Influence of age and gender on the 28-joint Disease Activity Score (DAS28) in rheumatoid arthritis**. *Annals of the rheumatic diseases* 2008, **67**(8):1127-1131.
35. Ranganath VK, Elashoff DA, Khanna D, Park G, Peter JB, Paulus HE: **Age adjustment corrects for apparent differences in erythrocyte sedimentation rate and C-reactive protein values at the onset of seropositive rheumatoid arthritis in younger and older patients**. *The Journal of rheumatology* 2005, **32**(6):1040-1042.
36. Kuznetsova A, Brockhoff PB, Christensen RHB: **Package 'lmertest'**. *R package version* 2015, **2**(0).

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

37. Yang XY, Di Zheng K, Lin K, Zheng G, Zou H, Wang JM, Lin YY, Chuka CM, Ge RS, Zhai W: **Energy metabolism disorder as a contributing factor of rheumatoid arthritis: a comparative proteomic and metabolomic study**. *PloS one* 2015, **10**(7).
38. Zhou J, Chan L, Zhou S: **Trigonelline: a plant alkaloid with therapeutic potential for diabetes and central nervous system disease**. *Current medicinal chemistry* 2012, **19**(21):3523-3531.
39. Zhou J-Y, Du X-H, Zhang Z, Qian G-S: **Trigonelline inhibits inflammation and protects β cells to prevent fetal growth restriction during pregnancy in a mouse model of diabetes**. *Pharmacology* 2017, **100**(5-6):209-217.
40. Foresti R, Green CJ, Motterlini R: **Generation of bile pigments by haem oxygenase: a refined cellular strategy in response to stressful insults**. In: *Biochemical Society Symposia: 2004*: Portland Press Limited; 2004: 177-192.
41. Bellner L, Vitto M, Patil KA, Dunn MW, Regan R, Laniado-Schwartzman M: **Exacerbated corneal inflammation and neovascularization in the HO-2 null mice is ameliorated by biliverdin**. *Experimental eye research* 2008, **87**(3):268-278.
42. Nakao A, Otterbein LE, Overhaus M, Sarady JK, Tsung A, Kimizuka K, Nalesnik MA, Kaizu T, Uchiyama T, Liu F: **Biliverdin protects the functional integrity of a transplanted syngeneic small bowel**. *Gastroenterology* 2004, **127**(2):595-606.
43. Sarady-Andrews JK, Liu F, Gallo D, Nakao A, Overhaus M, Ollinger R, Choi AM, Otterbein LE: **Biliverdin administration protects against endotoxin-induced acute lung injury in rats**. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2005, **289**(6):L1131-L1137.
44. Gibbs PE, Maines MD: **Biliverdin inhibits activation of NF-κB: Reversal of inhibition by human biliverdin reductase**. *International journal of cancer* 2007, **121**(11):2567-2574.
45. Öllinger R, Wang H, Yamashita K, Wegiel B, Thomas M, Margreiter R, Bach FH: **Therapeutic applications of bilirubin and biliverdin in transplantation**. *Antioxidants & redox signaling* 2007, **9**(12):2175-2186.
46. Peng Y-F, Wang J-L, Pan G-G: **The correlation of serum bilirubin levels with disease activity in patients with rheumatoid arthritis**. *Clinica Chimica Acta* 2017, **469**:187-190.
47. Fischman D, Valluri A, Gorrepati VS, Murphy ME, Peters I, Cheriyath P: **Bilirubin as a protective factor for rheumatoid arthritis: an NHANES study of 2003-2006 data**. *Journal of clinical medicine research* 2010, **2**(6):256.
48. Jangi S, Otterbein L, Robson S: **The molecular basis for the immunomodulatory activities of unconjugated bilirubin**. *The International Journal of Biochemistry & Cell Biology* 2013, **45**(12):2843-2851.
49. Enot DP, Lin W, Beckmann M, Parker D, Overy DP, Draper J: **Preprocessing, classification modeling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data**. *Nature Protocols* 2008, **3**(3):446-470.
50. Grissa D, Pétéra M, Brandolini M, Napoli A, Comte B, Pujos-Guillot E: **Feature selection methods for early predictive biomarker discovery using untargeted metabolomic data**. *Frontiers in molecular biosciences* 2016, **3**:30.
51. van Zanten JV, Ring C, Carroll D, Kitas GD: **Increased C reactive protein in response to acute stress in patients with rheumatoid arthritis**. *Annals of the rheumatic diseases* 2005, **64**(9):1299-1304.
52. Graf J, Scherzer R, Grunfeld C, Imboden J: **Levels of C-reactive protein associated with high and very high cardiovascular risk are prevalent in patients with rheumatoid arthritis**. *PLoS One* 2009, **4**(7):e6242.
53. Young SP, Kapoor SR, Viant MR, Byrne JJ, Filer A, Buckley CD, Kitas GD, Raza K: **The impact of inflammation on metabolomic profiles in patients with arthritis**. *Arthritis & Rheumatism* 2013, **65**(8):2015-2023.
54. Wells G, Becker J, Teng J, Dougados M, Schiff M, Smolen J, Aletaha D, Van Riel P: **Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid**

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. *Annals of the rheumatic diseases* 2009, **68**(6):954-960.

55. Patel CH, Leone RD, Horton MR, Powell JD: **Targeting metabolism to regulate immune responses in autoimmunity and cancer**. *Nature Reviews Drug Discovery* 2019, **18**(9):669-688.
56. Bettencourt IA, Powell JD: **Targeting metabolism as a novel therapeutic approach to autoimmunity, inflammation, and transplantation**. *The Journal of Immunology* 2017, **198**(3):999-1005.
57. Spies CM, Straub RH, Buttgereit F: **Energy metabolism and rheumatic diseases: from cell to organism**. *Arthritis research & therapy* 2012, **14**(3):216.
58. Yang Z, Matteson EL, Goronzy JJ, Weyand CM: **T-cell metabolism in autoimmune disease**. *Arthritis research & therapy* 2015, **17**(1):29.
59. Fearon U, Hanlon M, Wade S, Fletcher J: **Altered metabolic pathways regulate synovial inflammation in rheumatoid arthritis**. *Clinical & Experimental Immunology* 2019, **197**(2):170-180.
60. Chimenti M, Triggianese P, Conigliaro P, Candi E, Melino G, Perricone R: **The interplay between inflammation and metabolism in rheumatoid arthritis**. *Cell death & disease* 2015, **6**(9):e1887-e1887.
61. Sanchez-Lopez E, Cheng A, Guma M: **Can metabolic pathways be therapeutic targets in rheumatoid arthritis?** *Journal of clinical medicine* 2019, **8**(5):753.
62. Sung J, Wang Y, Chandrasekaran S, Witten DM, Price ND: **Molecular signatures from omics data: from chaos to consensus**. *Biotechnology journal* 2012, **7**(8):946-957.
63. Coras R, D Murillo-Saich J, Guma M: **Circulating Pro-and Anti-Inflammatory Metabolites and Its Potential Role in Rheumatoid Arthritis Pathogenesis**. *Cells* 2020, **9**(4):827.
64. Wegiel B, Otterbein L: **Go green: the anti-inflammatory effects of biliverdin reductase**. *Frontiers in pharmacology* 2012, **3**:47.
65. Wegiel B, Gallo D, Csizmadia E, Roger T, Kaczmarek E, Harris C, Zuckerbraun BS, Otterbein LE: **Biliverdin inhibits Toll-like receptor-4 (TLR4) expression through nitric oxide-dependent nuclear translocation of biliverdin reductase**. *Proceedings of the National Academy of Sciences* 2011, **108**(46):18849-18854.
66. Vogel ME, Zucker SD: **Bilirubin acts as an endogenous regulator of inflammation by disrupting adhesion molecule-mediated leukocyte migration**. *Inflammation and cell signaling* 2016, **3**(1).
67. Liu Y, Li P, Lu J, Xiong W, Oger J, Tetzlaff W, Cynader M: **Bilirubin possesses powerful immunomodulatory activity and suppresses experimental autoimmune encephalomyelitis**. *The Journal of Immunology* 2008, **181**(3):1887-1897.
68. Tang W, Wu Y, Wu B: **Heme catabolic pathway in inflammation and immune disorders**. *Frontiers in pharmacology* 2019, **10**:825.
69. Öllinger R, Yamashita K, Bilban M, Erat A, Kogler P, Thomas M, Csizmadia E, Usheva A, Margreiter R, Bach FH: **Bilirubin and biliverdin treatment of atherosclerotic diseases**. *Cell Cycle* 2007, **6**(1):39-43.
70. Pellock SJ, Redinbo MR: **Glucuronides in the gut: Sugar-driven symbioses between microbe and host**. *Journal of Biological Chemistry* 2017, **292**(21):8569-8576.
71. Takasuna K, Hagiwara T, Hirohashi M, Kato M, Nomura M, Nagai E, Yokoi T, Kamataki T: **Involvement of β-glucuronidase in intestinal microflora in the intestinal toxicity of the antitumor camptothecin derivative irinotecan hydrochloride (CPT-11) in rats**. *Cancer research* 1996, **56**(16):3752-3757.
72. LoGuidice A, Wallace BD, Bendel L, Redinbo MR, Boelsterli UA: **Pharmacologic targeting of bacterial β-glucuronidase alleviates nonsteroidal anti-inflammatory drug-induced enteropathy in mice**. *Journal of Pharmacology and Experimental Therapeutics* 2012, **341**(2):447-454.
73. Chen J, Wright K, Davis JM, Jeraldo P, Marietta EV, Murray J, Nelson H, Matteson EL, Taneja V: **An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis**. *Genome medicine* 2016, **8**(1):1-14.
74. He M, Harms AC, van Wijk E, Wang M, Berger R, Koval S, Hankemeier T, van der Greef J: **Role of amino acids in rheumatoid arthritis studied by metabolomics**. *International Journal of Rheumatic Diseases* 2019, **22**(1):38-46.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20193664; this version posted March 25, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

75. Tilvawala R, Nguyen SH, Maurais AJ, Nemmara VV, Nagar M, Salinger AJ, Nagpal S, Weerapana E, Thompson PR: **The rheumatoid arthritis-associated citrullinome**. *Cell chemical biology* 2018, **25**(6):691-704. e696.
76. Marietta EV, Murray JA, Luckey DH, Jeraldo PR, Lamba A, Patel R, Luthra HS, Mangalam A, Taneja V: **Suppression of inflammatory arthritis by human gut-derived Prevotella histicola in humanized mice**. *Arthritis & rheumatology* 2016, **68**(12):2878-2888.
77. Liu X, Zou Q, Zeng B, Fang Y, Wei H: **Analysis of fecal Lactobacillus community structure in patients with early rheumatoid arthritis**. *Current microbiology* 2013, **67**(2):170-176.
78. Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, Wu X, Li J, Tang L, Li Y: **The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment**. *Nature medicine* 2015, **21**(8):895-905.
79. Sung J, Kim S, Cabatbat JJT, Jang S, Jin Y-S, Jung GY, Chia N, Kim P-J: **Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis**. *Nature communications* 2017, **8**(1):1-12.
80. Wilmanski T, Rappaport N, Earls JC, Magis AT, Manor O, Lovejoy J, Omenn GS, Hood L, Gibbons SM, Price ND: **Blood metabolome predicts gut microbiome α-diversity in humans**. *Nature biotechnology* 2019, **37**(10):1217-1228.
81. Gupta VK, Kim M, Bakshi U, Cunningham KY, Davis JM, Lazaridis KN, Nelson H, Chia N, Sung J: **A predictive index for health status using species-level gut microbiome profiling**. *Nature communications* 2020, **11**(1):1-16.