

## **Differences in the serum metabolome and lipidome identify potential biomarkers for seronegative rheumatoid arthritis versus psoriatic arthritis**

M Margarida Souto-Carneiro, PhD; Lilla Tóth; Rouven Behnisch, MSc; Konstantin Urbach, MD; Karel D. Klika, PhD; Rui A Carvalho, PhD; Hanns-Martin Lorenz, MD

### **SUPPLEMENTARY MATERIALS**

#### **Patients and Methods**

##### **Patient selection**

Disease activity for both groups was assessed at the time of blood collection through quantification of tender and swollen joint counts, erythrocyte sedimentation rate, C-reactive protein levels and the visual analog scale-grading of the patient. Treatment modalities were registered in the data sheets. Disease activity groups within both negRA and PsA patients were defined according to the DAS28-CRP (4 variables) score(1, 2) as remission ( $<2.6$ ), low disease activity ( $\geq 2.6 < 3.2$ ) and high disease activity ( $>3.2$ ). Patients were excluded from the study whenever any of the following criteria were met: 1) daily corticosteroid intake  $>20$  mg; 2) presence of type I or II diabetes; 3) less than 12 months after finishing antitumor therapy; 4) pregnancy; 5) active viral and/or bacterial infection; 6) presence of other chronic arthralgias or 7) denial of consent to donate blood for the study.

##### **Nuclear Magnetic Resonance (NMR) Spectroscopic Analysis**

###### ***Sample acquisition***

Serum was obtained by centrifugation of venous blood and immediately stored at  $-20^{\circ}\text{C}$  until analysis. For preparing samples for  $^1\text{H}$  NMR spectroscopy, 140  $\mu\text{L}$  of each serum was mixed with 35  $\mu\text{L}$  of a standard solution of phosphate buffer (0.2 M) and sodium fumarate (10 mM) in  $\text{D}_2\text{O}$  (99.9%). Samples with a total volume of 175  $\mu\text{L}$  were transferred into 2.5 mm NMR tubes and  $^1\text{H}$  single-pulse (sw = 10 kHz, ns = 32, td = 60k, aq = 3 s) and Carr–Purcell–Meiboom–Gill (CPMG) (sw = 10 kHz, ns = 512, spin–spin relaxation delay = 162.4 ms, spin–spin echo delay = 400  $\mu\text{s}$ , td = 60k, aq = 3 s)(3) NMR-spectra were acquired at  $25^{\circ}\text{C}$  using a 600 MHz Bruker NMR spectrometer equipped with a 5-mm inverse-configuration probe.

###### ***Data processing***

NMR spectra were processed using TopSpin 3.6.2 software (Bruker BioSpin GmbH). Prior to Fourier transformation of the raw data, a 0.2 Hz exponential decay function was applied. Phasing, baseline correction and spectral calibration were executed manually.

For the  $^1\text{H}$  single-pulse NMR-spectra, 13 variable-sized buckets from 0–8.50 parts per million (ppm) were evaluated excluding the region of the solvent signal (Regions A to L, Figure 1A) while the  $^1\text{H}$  CPMG NMR-spectra were divided into 10 variable-sized buckets from 0.88–8.55 ppm excluding regions A and C of the single-pulse spectra and the region of the solvent signal (Figure 1B). Binning was carried out using Amix 4.0.1 software (Bruker BioSpin GmbH). Matrices were created from the binned data taken from the  $^1\text{H}$  single-pulse and CPMG NMR-spectra in which each row corresponded to an individual sample.

The associated spectral regions for the single identified metabolites were integrated for quantification using AMIX software. Absolute quantification of the metabolite concentrations was achieved by comparing the integral of a signal from each metabolite relative to the signal integral of the internal fumarate standard.

### Statistical Analysis

Multivariate statistical analysis was performed using MetaboAnalyst 4.0 software(4) on the binned data matrixes. After normalization to a constant sum and auto-scaling, chemometrics analysis was done using partial least squares–discriminant analysis (PLS-DA) with a 10-fold cross-validation and classification and feature selection analysis was performed using a Random Forest algorithm with 500 trees, 7 predictors and with randomness.

The D’Agostino–Pearson omnibus normality test was used to verify whether the variables followed a normal distribution. The for normally distributed variables an unpaired t-test with Welch’s correction was used to determine whether the concentration of each metabolite differed between disease groups, for those not following a normal distribution the comparison was made using the non-parametric Mann-Whitney test. Correlation coefficients between clinical/demographic variables and metabolites or lipid concentrations were calculated using the parametric Pearson product-moment correlation coefficient ( $R$ ). Correlations were considered weak for  $0.1 < |R| \leq 0.3$ , moderate for  $0.3 < |R| \leq 0.5$ , strong for  $0.5 < |R| \leq 0.7$  and very strong for  $|R| > 0.7$ . For all tests, p-values  $< 0.05$  were considered statistically significant. Univariate statistical analyses were calculated using either GraphPad Prism or MedCalc software.

Multivariate analysis of association (MANOVA) was carried using SPSS (Build 1.0.0.1327, IBM Corp.) to assess the combined impact of all possible confounders on the distribution of the different single metabolites between groups: age, disease duration, therapeutic regimen, disease activity (DAS28), body-mass index (BMI), C-reactive protein (CRP) positivity and gender. We assumed a heterogenous distribution of variance, thus used the Wilks' Lambda to determine the statistical significance(5). To account for multiple ANOVA in defining which dependent variables present differences for the independent variables we considered an alpha-value of 0.025.

To calculate the one-way and multi-way MANOVA the continuous variables age, disease duration, BMI, and DAS28 were converted into nominal classes:

Age (arbitrary): A1-  $\geq 30y < 45y$ ; A2-  $\geq 45y < 60y$ ; A3-  $\geq 60y < 75y$ ; A4-  $\geq 75y$

Disease Duration (arbitrary): Short-  $< 5y$ ; Medium-  $\geq 5y < 10y$ ; Long-  $\geq 10y < 20y$ ; Very Long-  $\geq 20y$

DAS28 (according to (1, 2): Remission-  $< 2.6$ ; Low Activity-  $\geq 2.6 < 3.2$ ; High Activity-  $\geq 3.2$

BMI (according to the WHO classification): Normal-  $\geq 18.5 \leq 24.9$ ; Overweight-  $\geq 25 \leq 29.9$ ; Obese-  $\geq 30$ .

CRP (according to (6)): Negative-  $< 2.0 \text{ mg/L}$ ; Positive-  $\geq 2 \text{ mg/L}$

### **Pathway analysis**

For visualization of metabolic pathway differences between groups, all metabolites used in the targeted analysis were mapped using the Metabolite Set Enrichment Analysis tool from MetaboAnalyst. The quantitative enrichment analysis (QEA) was performed based on the compound concentrations. After assignment of the chemical similarity ID (PubChem ID) and biochemical pathway ID (KEGG ID) the October 2019 KEGG metabolic pathway associated metabolite sets was chosen to build a summary plot of the QEA.

### **Diagnostic Model**

#### ***Variable selection***

The diagnostic model was developed using a multivariate logistic regression model. Model selection was carried out by an automatic forward-backward stepwise selection algorithm that

started with the full model with all covariates and in each step excluded the variable with the highest p-value  $> 0.05$  (backwards selection) and then searched through the excluded variables for a covariate that improved the model having a p-value  $< 0.049$  (forward selection).

Two additional machine learning classification methods were tested, the random forest algorithm and the naive Bayes classifier. However, both yielded models with lower accuracy.

### ***Model Validation***

To evaluate the model a 10-fold cross-validation (CV) was performed, i.e. the sample was randomly split into 10 equally sized subsamples and the model was fitted to 9 subsamples and validated on the remaining one. This was repeated 10 times with each of the subsamples used exactly once as the validation data. This procedure ensures that the model is not evaluated on the same data used for model fitting and hence it avoids overfitting. For each of these 10 analyses, the area under the curve (AUC) was computed. The mean value of these 10 AUC values gives an estimate for the “true” AUC of the underlying logistic regression model.

A further model validation was done with a blinded external cohort to evaluate the predictive power using ROC analysis.

### **Supplementary Figure Legends:**

**Supplementary Figure 1: Influence of demographic and clinical variables in the metabolome of PsA and negRA samples.** Scores plot of the principal components 1 and 2 of the partial least squares discriminant analyses and out-of-bag error rates plots of the random forest classification models to predict (A) demographic (age, BMI, gender), and (B) clinical (CRP positivity, DAS28, Disease Duration (Dis\_Dur) and therapy) based on the metabolomic profile represented by the region A through L of the  $^1\text{H}$  NMR spectra.

### **Supplementary Tables Index:**

ST1- Detailed results of the one and multi-way MANOVA tests for clinical covariates in the discovery cohort.

ST2- Test of between-subjects effects for the MANOVA analyses for clinical covariates in the discovery cohort.

ST3- Detailed results of the one and multi-way MANOVA tests for demographic covariates in the discovery cohort.

ST4- Test of between-subjects effects for the MANOVA analyses for demographic covariates in the discovery cohort.

ST5- Mean metabolite concentration in each patient group of the discovery cohort.

ST6- Univariate ROC curve analysis.

ST7- Clinical, demographic and model classification data of the blinded validation cohort.

## References

1. Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* 2010;62:2569-81.
2. van der Heijde DM, van 't Hof MA, van Riel PL, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 1990;49:916-20.
3. Meiboom S, Gill D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments* 1958;29:688-91.
4. Chong J, Wishart DS, Xia J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics* 2019;68:e86.
5. Ates C, Kaymaz O, Kale HE, et al. Comparison of Test Statistics of Nonnormal and Unbalanced Samples for Multivariate Analysis of Variance in terms of Type-I Error Rates. *Comput Math Methods Med* 2019;2019:2173638.
6. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *J Clin Invest* 2003;111:1805-12.

**Supplementary Table ST1-** Detailed results of the one and multi-way MANOVA tests for clinical covariates (CRP status; DAS28 score; Disease Duration and Therapy). Significance is indicated by Wilks' lambda for  $p < 0.025$  in the discovery cohort.

Effect	Value	F	Hypothesis degrees of freedom	Error degrees of freedom	Significance (p-value)	Partial Eta Squared	Noncent. Parameter	Observed Power
Disease	0.299	2.187	30	28	<b>0.020</b>	0.701	65.6	0.905
CRP	0.583	0.666	30	28	0.861	0.417	20	0.261
DAS28	0.114	1.836	60	56	<b>0.011</b>	0.663	110.2	0.988
Disease Duration	0.055	1.543	90	84.7	<b>0.022</b>	0.620	138.3	0.996
Therapy	0.136	1.598	60	56	0.039	0.631	95.9	0.968
Disease + CRP	0.319	1.995	30	28	0.035	0.681	59.8	0.865
Disease + DAS28	0.224	1.039	60	56	0.444	0.527	62.3	0.785
Disease + Disease Duration	0.109	1.032	90	84.7	0.442	0.522	92.6	0.924
Disease + Therapy	0.153	1.453	60	56	0.080	0.609	87.2	0.945
Disease + CRP+DAS28	0.583	0.669	30	28	0.859	0.417	20.1	0.263
Disease + CRP+Disease Duration	0.221	1.053	60	56	0.424	0.530	63.2	0.793
Disease + CRP+Therapy	0.575	0.690	30	28	0.840	0.425	20.7	0.274
Disease + DAS28+Disease Duration	0.165	0.777	90	84.7	0.880	0.452	69.7	0.772
Disease + DAS28+Therapy	0.670	0.460	30	28	0.980	0.330	13.8	0.161
Disease + Disease Duration + Therapy	0.198	1.166	60	56	0.282	0.555	69.9	0.852

**Supplementary Table ST2-** Test of between-subjects effects for the MANOVA analyses for clinical covariates with a significance of  $p < 0.025$  in the discovery cohort.

Independent Variable	Dependent Variable	Type III Sum of Squares	Degrees of Freedom	Mean Square	F	Significance (p-value)	Partial Eta Squared	Noncent0. Parameter	Observed Power <sup>ae</sup>
Disease	Acetate	0.094	1	0.094	2.264	0.138	0.038	2.264	0.221
	Acetoacetate	0.050	1	0.050	3.349	0.072	0.055	3.349	0.326
	Alanine	0.903	1	0.903	0.806	0.373	0.014	0.806	0.087
	$\beta$ -hydroxybutyrate	4.140E-5	1	4.140E-5	0.010	0.921	0.000	0.010	0.026
	Choline	0.002	1	0.002	4.427	0.040	0.072	4.427	0.427
	Citrate	0.019	1	0.019	0.423	0.518	0.007	0.423	0.056
	Creatine	0.016	1	0.016	1.717	0.195	0.029	1.717	0.169
	Creatine-P	0.005	1	0.005	0.910	0.344	0.016	0.910	0.096
	Creatinine	0.001	1	0.001	0.633	0.429	0.011	0.633	0.073
	Formate	6.356E-5	1	6.356E-5	0.545	0.464	0.009	0.545	0.066
	Glucose	8.016	1	8.016	0.152	0.698	0.003	0.152	0.036
	Glutamate	0.642	1	0.642	3.680	0.060	0.061	3.680	0.357
	Glutamine	0.091	1	0.091	4.604	0.036	0.075	4.604	0.443
	Histidine	2.617E-5	1	2.617E-5	0.001	0.975	0.000	0.001	0.025
	Isoleucine	0.007	1	0.007	2.032	0.159	0.034	2.032	0.199
	L2/L1	0.201	1	0.201	0.471	0.495	0.008	0.471	0.060
	L3/L1	0.012	1	0.012	1.729	0.194	0.029	1.729	0.170
	L4/L1	0.001	1	0.001	0.304	0.583	0.005	0.304	0.047
L5/L1	0.005	1	0.005	1.889	0.175	0.032	1.889	0.185	
L6/L1	0.013	1	0.013	7.292	0.009	0.113	7.292	0.655	

	L7/L1	0.002	1	0.002	0.465	0.498	0.008	0.465	0.060
	Lactate	19.951	1	19.951	0.884	0.351	0.015	0.884	0.094
	Leucine	0.074	1	0.074	4.815	0.032	0.078	4.815	0.462
	Phenylalanine	0.002	1	0.002	2.660	0.108	0.045	2.660	0.259
	Proline	3.713	1	3.713	3.701	0.059	0.061	3.701	0.359
	Sarcosine	0.011	1	0.011	4.404	0.040	0.072	4.404	0.425
	Succinate	0.001	1	0.001	2.539	0.117	0.043	2.539	0.247
	Threonine	0.010	1	0.010	0.235	0.630	0.004	0.235	0.042
	Tyrosine	0.002	1	0.002	0.232	0.632	0.004	0.232	0.042
	Valine	0.658	1	0.658	3.551	0.065	0.059	3.551	0.345
<b>DAS28</b>	Acetate	0.004	2	0.002	0.044	0.957	0.002	0.087	0.029
	Acetoacetate	0.069	2	0.035	2.337	0.106	0.076	4.674	0.342
	Alanine	0.173	2	0.086	0.077	0.926	0.003	0.154	0.032
	$\beta$ -hydroxybutyrate	0.001	2	0	0.070	0.933	0.002	0.140	0.031
	Choline	0.003	2	0.002	4.947	0.010	0.148	9.894	0.694
	Citrate	0.023	2	0.012	0.263	0.770	0.009	0.526	0.050
	Creatine	0.015	2	0.008	0.821	0.445	0.028	1.643	0.116
	Creatine-P	0.007	2	0.004	0.675	0.513	0.023	1.351	0.097
	Creatinine	0.003	2	0.002	1.118	0.334	0.038	2.236	0.156
	Formate	0.001	2	0	2.546	0.087	0.082	5.092	0.374
	Glucose	222.839	2	111.420	2.119	0.130	0.069	4.237	0.308
	Glutamate	0.709	2	0.355	2.032	0.140	0.067	4.065	0.294
	Glutamine	0.054	2	0.027	1.351	0.267	0.045	2.702	0.190
	Histidine	0.014	2	0.007	0.249	0.780	0.009	0.499	0.049
	Isoleucine	0.000	2	0	0.060	0.942	0.002	0.121	0.030
	L2/L1	3.395	2	1.698	3.974	0.024	0.122	7.948	0.580
	L3/L1	0.015	2	0.008	1.120	0.333	0.038	2.240	0.157



	L4/L1	0.018	2	0.009	2.912	0.063	0.093	5.824	0.430
	L5/L1	0.014	2	0.007	2.907	0.063	0.093	5.814	0.429
	L6/L1	0.002	2	0.001	0.695	0.503	0.024	1.390	0.100
	L7/L1	0.040	2	0.020	4.163	0.021	0.127	8.325	0.604
	Lactate	16.215	2	8.107	0.359	0.700	0.012	0.719	0.060
	Leucine	0.005	2	0.002	0.147	0.863	0.005	0.295	0.038
	Phenylalanine	0.002	2	0.001	1.115	0.335	0.038	2.230	0.156
	Proline	4.374	2	2.187	2.180	0.122	0.071	4.359	0.317
	Sarcosine	0.003	2	0.002	0.652	0.525	0.022	1.305	0.094
	Succinate	0.000	2	0	0.443	0.644	0.015	0.886	0.070
	Threonine	0.140	2	0.070	1.684	0.195	0.056	3.368	0.240
	Tyrosine	0.002	2	0.001	0.122	0.886	0.004	0.244	0.036
	Valine	0.303	2	0.151	0.817	0.447	0.028	1.634	0.115
<b>Disease Duration</b>	Acetate	0.094	3	0.031	0.754	0.525	0.038	2.261	0.128
	Acetoacetate	0.126	3	0.042	2.826	0.047	0.129	8.477	0.534
	Alanine	10.678	3	3.559	3.178	0.031	0.143	9.534	0.597
	$\beta$ -hydroxybutyrate	0.012	3	0.004	0.996	0.402	0.050	2.987	0.172
	Choline	0.001	3	0	0.835	0.480	0.042	2.506	0.143
	Citrate	1.060	3	0.353	8.051	<0.0001	0.298	24.154	0.973
	Creatine	0.044	3	0.015	1.598	0.200	0.078	4.793	0.290
	Creatine-P	0.133	3	0.044	8.380	<0.0001	0.306	25.139	0.978
	Creatinine	0.002	3	0.001	0.441	0.724	0.023	1.324	0.079
	Formate	0.001	3	0	1.952	0.131	0.093	5.857	0.363
	Glucose	1181.733	3	393.911	7.490	<0.0001	0.283	22.469	0.961
	Glutamate	0.725	3	0.242	1.386	0.256	0.068	4.157	0.247
	Glutamine	0.073	3	0.024	1.227	0.308	0.061	3.682	0.216
	Histidine	0.318	3	0.106	3.906	0.013	0.171	11.719	0.708

Isoleucine	0.011	3	0.004	0.966	0.415	0.048	2.898	0.166
L2/L1	0.760	3	0.253	0.593	0.622	0.030	1.778	0.102
L3/L1	0.023	3	0.008	1.121	0.348	0.056	3.362	0.195
L4/L1	0.013	3	0.004	1.377	0.259	0.068	4.131	0.245
L5/L1	0.002	3	0.001	0.336	0.799	0.017	1.009	0.064
L6/L1	0.010	3	0.003	1.967	0.129	0.094	5.902	0.366
L7/L1	0.008	3	0.003	0.548	0.651	0.028	1.644	0.095
Lactate	16.116	3	5.372	0.238	0.869	0.012	0.714	0.052
Leucine	0.081	3	0.027	1.752	0.167	0.084	5.256	0.322
Phenylalanine	0.002	3	0.001	0.976	0.411	0.049	2.927	0.168
Proline	0.162	3	0.054	0.054	0.983	0.003	0.162	0.030
Sarcosine	0.025	3	0.008	3.206	0.030	0.144	9.618	0.601
Succinate	0.002	3	0.001	1.546	0.212	0.075	4.638	0.279
Threonine	0.308	3	0.103	2.464	0.072	0.115	7.392	0.465
Tyrosine	0.120	3	0.040	5.976	0.001	0.239	17.929	0.902
Valine	2.651	3	0.884	4.767	0.005	0.201	14.301	0.809

**Supplementary Table ST3-** Detailed results of the one and multi-way MANOVA tests for demographic covariates (Age; BMI and Gender). Significance is indicated by Wilks' lambda for  $p < 0.025$  in the discovery cohort.

Effect	Value	F	Hypothesis degrees of freedom	Error degrees of freedom	Significance (p-value)	Partial Eta Squared	Noncent. Parameter	Observed Power
<b>Age</b>	0.432	0.628	90	174.5	0.993	0.244	56.364	0.775
<b>Gender</b>	0.692	0.862	30	58	0.665	0.308	25.858	0.521
<b>BMI</b>	0.568	0.631	60	116	0.975	0.246	37.856	0.606
<b>Disease + Age</b>	0.174	1.541	90	174.5	<b>0.008</b>	0.442	138.270	1.000
<b>Disease + Gender</b>	0.714	0.775	30	58	0.773	0.286	23.258	0.457
<b>Disease + BMI</b>	0.334	1.412	60	116	0.057	0.422	84.704	0.987
<b>Disease + Age + BMI</b>	0.188	1.450	90	174.5	<b>0.019</b>	0.427	130.062	0.999
<b>Disease + Gender + BMI</b>	0.394	1.145	60	116	0.264	0.372	68.729	0.947

**Supplementary Table ST4-** Test of between-subjects effects for the MANOVA analyses for demographic covariates with a significance of  $p < 0.025$  in the discovery cohort.

Independent Variable	Dependent Variable	Type III Sum of Squares	Degrees of Freedom	Mean Square	F	Significance (p-value)	Partial Eta Squared	Noncent. Parameter	Observed Power
Disease + Age	Acetate	0.024	3	0.008	0.231	0.874	0.008	0.694	0.052
	Acetoacetate	0.018	3	0.006	0.348	0.791	0.012	1.043	0.067
	Alanine	2.554	3	0.851	0.801	0.496	0.027	2.404	0.141
	$\beta$ -hydroxybutyrate	0.019	3	0.006	1.619	0.191	0.053	4.857	0.304
	Choline	0.001	3	0	0.869	0.461	0.029	2.606	0.153
	Citrate	0.400	3	0.133	3.415	0.021	0.105	10.244	0.651
	Creatine	0.068	3	0.023	2.748	0.048	0.087	8.244	0.534
	Creatine-P	0.030	3	0.010	1.279	0.287	0.042	3.836	0.233
	Creatinine	0	3	0	0.111	0.953	0.004	0.334	0.037
	Formate	0.001	3	0	1.338	0.267	0.044	4.013	0.245
	Glucose	187.197	3	62.399	1.294	0.282	0.043	3.882	0.236
	Glutamate	1.629	3	0.543	3.790	0.013	0.116	11.369	0.707
	Glutamine	0.158	3	0.053	2.955	0.037	0.092	8.865	0.573
	Histidine	0.060	3	0.020	0.888	0.451	0.030	2.664	0.156
	Isoleucine	0.011	3	0.004	1.126	0.343	0.037	3.379	0.202
	L2/L1	3.260	3	1.087	3.392	0.022	0.105	10.177	0.647
	L3/L1	0.001	3	0	0.045	0.987	0.002	0.135	0.030
	L4/L1	0.008	3	0.003	0.868	0.461	0.029	2.603	0.153
	L5/L1	0.003	3	0.001	0.469	0.704	0.016	1.408	0.085
	L6/L1	0.001	3	0	0.239	0.869	0.008	0.718	0.053
L7/L1	0.034	3	0.011	2.656	0.053	0.084	7.969	0.517	
Lactate	21.364	3	7.121	0.374	0.772	0.013	1.122	0.071	
Leucine	0.043	3	0.014	0.934	0.428	0.031	2.802	0.165	

	Phenylalanine	0.003	3	0.001	1.326	0.271	0.044	3.977	0.242
	Proline	0.752	3	0.251	0.270	0.847	0.009	0.811	0.057
	Sarcosine	0.007	3	0.002	0.539	0.657	0.018	1.618	0.096
	Succinate	0.001	3	0	1.150	0.334	0.038	3.450	0.207
	Threonine	0.073	3	0.024	0.628	0.599	0.021	1.883	0.110
	Tyrosine	0.036	3	0.012	2.112	0.105	0.068	6.336	0.407
	Valine	0.508	3	0.169	0.875	0.457	0.029	2.626	0.154
<b>Disease + Age + BMI</b>	Acetate	0.028	3	0.009	0.268	0.848	0.009	0.805	0.056
	Acetoacetate	0.043	3	0.014	0.840	0.475	0.028	2.521	0.148
	Alanine	3.337	3	1.112	1.047	0.376	0.035	3.141	0.187
	$\beta$ -hydroxybutyrate	0.004	3	0.001	0.374	0.772	0.013	1.123	0.071
	Choline	0.003	3	0.001	2.277	0.085	0.073	6.831	0.441
	Citrate	0.278	3	0.093	2.373	0.076	0.076	7.120	0.461
	Creatine	0.019	3	0.006	0.787	0.504	0.026	2.361	0.138
	Creatine-P	0.031	3	0.010	1.285	0.285	0.042	3.855	0.234
	Creatinine	0.004	3	0.001	0.903	0.443	0.030	2.709	0.159
	Formate	0.001	3	0	1.493	0.222	0.049	4.479	0.277
	Glucose	138.723	3	46.241	0.959	0.416	0.032	2.877	0.170
	Glutamate	0.431	3	0.144	1.002	0.396	0.033	3.005	0.178
	Glutamine	0.043	3	0.014	0.802	0.496	0.027	2.406	0.141
	Histidine	0.163	3	0.054	2.413	0.072	0.077	7.239	0.469
	Isoleucine	0.007	3	0.002	0.673	0.571	0.023	2.018	0.118
	L2/L1	0.515	3	0.172	0.536	0.659	0.018	1.608	0.096
	L3/L1	0.022	3	0.007	1.117	0.347	0.037	3.352	0.200
	L4/L1	0.007	3	0.002	0.753	0.523	0.025	2.260	0.132
	L5/L1	0.005	3	0.002	0.802	0.496	0.027	2.405	0.141
	L6/L1	0.002	3	0.001	0.485	0.694	0.016	1.455	0.088

L7/L1	0.003	3	0.001	0.208	0.890	0.007	0.625	0.049
Lactate	14.262	3	4.754	0.250	0.861	0.009	0.749	0.054
Leucine	0.041	3	0.014	0.889	0.450	0.030	2.667	0.157
Phenylalanine	0.004	3	0.001	1.653	0.183	0.054	4.958	0.311
Proline	1.531	3	0.510	0.550	0.650	0.019	1.650	0.098
Sarcosine	0.036	3	0.012	2.599	0.057	0.082	7.797	0.506
Succinate	0.002	3	0.001	1.348	0.264	0.044	4.044	0.247
Threonine	0.152	3	0.051	1.301	0.279	0.043	3.903	0.237
Tyrosine	0.024	3	0.008	1.444	0.236	0.047	4.331	0.267
Valine	0.539	3	0.180	0.929	0.430	0.031	2.786	0.164

**Supplementary Table ST5-** Mean ( $\pm$  SEM) metabolite concentration in each patient group of the discovery cohort.

Variable	Group	Mean	Std. Error Mean	p-value
L7/L1	RA	0.1892	0.00823	0.623
	PSA	0.1950	0.00827	
L6/L1	RA	0.1445	0.00447	0.014
	PSA	0.1620	0.00538	
L5/L1	RA	0.1369	0.00658	0.008
	PSA	0.1607	0.00587	
L4/L1	RA	0.2947	0.00670	0.594
	PSA	0.3000	0.00734	
L3/L1	RA	0.3769	0.01014	0.033
	PSA	0.4073	0.00979	
L2/L1	RA	2.0691	0.06950	0.479
	PSA	2.1440	0.07939	
Glucose (mM)	RA	13.1760	0.87117	0.310
	PSA	14.5342	1.00636	
Acetoacetate (mM)	RA	0.1900	0.01982	0.110
	PSA	0.2301	0.01493	
Alanine (mM)	RA	1.3597	0.06633	0.002
	PSA	1.8835	0.15225	
Acetate (mM)	RA	0.1374	0.01070	0.014
	PSA	0.2040	0.02448	
Succinate (mM)	RA	0.0495	0.00256	0.053
	PSA	0.0425	0.00254	
Citrate (mM)	RA	0.3638	0.02423	0.539
	PSA	0.3882	0.03151	
Histidine (mM)	RA	0.2862	0.01274	0.125
	PSA	0.3274	0.02342	
Tyrosine (mM)	RA	0.1407	0.00808	0.165
	PSA	0.1601	0.01127	
Threonine (mM)	RA	0.4336	0.02419	0.006
	PSA	0.5363	0.02761	
Creatinine (mM)	RA	0.0385	0.00299	0.073
	PSA	0.0491	0.00506	
Creatine-P (mM)	RA	0.1645	0.01357	0.759
	PSA	0.1698	0.01039	
Creatine (mM)	RA	0.1100	0.01178	0.048
	PSA	0.1429	0.01150	
Sarcosine (mM)	RA	0.0593	0.00678	0.505
	PSA	0.0670	0.00927	

<b>β-hydroxybutyrate (mM)</b>	RA	0.0961	0.00871	0.888
	PSA	0.0945	0.00742	
<b>Valine (mM)</b>	RA	0.7574	0.05038	0.003
	PSA	0.9976	0.06045	
<b>Lactate (mM)</b>	RA	2.7169	0.18548	0.029
	PSA	4.0829	0.58684	
<b>Leucine (mM)</b>	RA	0.4431	0.01462	0.043
	PSA	0.4886	0.01676	
<b>Isoleucine (mM)</b>	RA	0.1782	0.00755	0.351
	PSA	0.1881	0.00742	
<b>Glutamate (mM)</b>	RA	2.1587	0.04929	0.944
	PSA	2.1638	0.05327	
<b>Glutamine (mM)</b>	RA	0.5505	0.01960	0.939
	PSA	0.5525	0.01800	
<b>Choline (mM)</b>	RA	0.0631	0.00235	0.007
	PSA	0.0738	0.00309	
<b>Proline (mM)</b>	RA	1.6450	0.04578	0.263
	PSA	1.8045	0.13408	
<b>Phenylalanine (mM)</b>	RA	0.0832	0.00364	0.006
	PSA	0.0689	0.00361	
<b>Formate (mM)</b>	RA	0.0278	0.00150	0.488
	PSA	0.0263	0.00165	



**Supplementary Table ST6-** Univariate ROC curve analysis (non-parametric assumption) for the discovery cohort.

Variable	AUC	Std. Error	Asymptotic p-value	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
L7/L1	0.508	0.053	0.882	0.405	0.611
L6/L1	0.614	0.051	0.034	0.513	0.714
L5/L1	0.623	0.051	0.022	0.523	0.722
L4/L1	0.519	0.053	0.728	0.415	0.622
L3/L1	0.586	0.052	0.109	0.483	0.688
L2/L1	0.490	0.052	0.853	0.388	0.592
Glucose	0.538	0.053	0.482	0.434	0.641
Acetoacetate	0.620	0.052	0.026	0.517	0.722
Alanine	0.648	0.049	0.006	0.551	0.745
Acetate	0.623	0.051	0.022	0.523	0.722
Succinate	0.354	0.050	0.007	0.257	0.452
Citrate	0.480	0.052	0.705	0.377	0.583
Histidine	0.489	0.052	0.841	0.387	0.591
Tyrosine	0.520	0.052	0.713	0.417	0.622
Threonine	0.630	0.051	0.015	0.530	0.729
Creatinine	0.541	0.052	0.441	0.438	0.644
Creatine-P	0.545	0.055	0.399	0.438	0.652
Creatine	0.615	0.051	0.032	0.514	0.716
Sarcosine	0.487	0.053	0.812	0.383	0.591
$\beta$ -hydroxybutyrate	0.483	0.053	0.752	0.379	0.587
Valine	0.680	0.049	0.001	0.584	0.777
Lactate	0.566	0.052	0.217	0.465	0.667
Leucine	0.598	0.051	0.067	0.498	0.698
Isoleucine	0.547	0.053	0.379	0.444	0.651
Glutamate	0.499	0.053	0.981	0.396	0.602
Glutamine	0.496	0.054	0.940	0.390	0.602
Choline	0.611	0.052	0.038	0.510	0.712
Proline	0.430	0.052	0.189	0.328	0.531
Phenylalanine	0.361	0.049	0.009	0.264	0.458
Formate	0.427	0.054	0.170	0.321	0.532

AUC- area under the curve

**Supplementary Table ST7-** Clinical, demographic and model classification data of the separate blinded validation cohort. Underlined samples indicate the individuals with unclear clinical parameters after re-evaluation.

Sample	Gender	Age (years)	Diagnosis	Disease Length (years)	DAS28	CRP (mg/L)	GC (mg)	MTX (mg)	Leflunomid (mg)	Biologics	Probability PsA	Probability RA	Model Diagnosis
VB 1	female	50	negRA	8	1.15	1.9	0	10	0	--	0.131	0.869	negRA
VB 2	female	60	negRA	2	3.15	3.8	5	15	20	--	0.08	0.92	negRA
VB 3	female	34	PSA	0	4.74	98.6	5	0	0	--	0.464	0.536	negRA
VB 4	female	67	PSA	1	2.53	5.8	0	0	0	yes	0.25	0.75	negRA
VB 5	male	53	negRA	3	1.15	1.9	5	20	0	--	0.961	0.039	PSA
VB 6	male	60	PSA	1	1.15	1.9	10	15	0	--	0.527	0.473	PSA
VB 7	female	69	negRA	18	4.02	9.2	0	0	0	--	0.092	0.908	negRA
VB 8	female	52	negRA	1	3.69	1.9	20	0	0	--	0.063	0.937	negRA
VB 9	female	68	negRA	6	4.03	10.7	10	10	0	--	0.027	0.973	negRA
VB 10	male	54	PSA	7	1.15	1.9	5	0	10	--	0.93	0.07	PSA
VB 11	male	43	PSA	0	3.81	6.8	0	0	0	--	0.458	0.542	negRA
VB 12	male	76	negRA	2	3.79	34.1	10		20	--	0.024	0.976	negRA
VB 13	female	58	negRA	27	3.72	4.6	5	10	0	yes	0.041	0.959	negRA
<u>VB 14</u>	male	62	mixed	2	2.81	12.8	5	20	0	--	0.279	0.721	negRA
VB 15	male	40	PSA	2	3.96	1.9	0	10	0	--	0.546	0.454	PSA
VB 16	male	71	PSA	9	1.79	4	5	0	0	--	0.67	0.33	PSA
VB 17	female	55	PSA	2	2.37	21	0	0	0	--	0.023	0.977	negRA
VB 18	female	44	negRA	3	2.22	1.9	0	15	10	--	0.765	0.235	PSA
VB 19	male	53	PSA	27	1.67	2.7	12.5	10	0	--	0.758	0.242	PSA
VB 20	male	58	PSA	10	1.82	4.4	4	20	0	yes	0.917	0.083	PSA
VB 21	male	61	negRA	27	1.15	1.9	5	15	0	--	0.187	0.813	negRA
<u>VB 22</u>	female	50	mixed	5	1.62	2.3	5	15	0	--	0.593	0.407	PSA
<u>VB 23</u>	male	42	mixed	10	1.3	1.9	7.5	0	20	yes	0.83	0.17	PSA
VB 24	female	32	PSA	3	3.99	11.3	7.5	0	0	--	0.656	0.344	PSA
VB 25	male	30	PSA	0	3.37	7.4	0	0	0	yes	0.929	0.071	PSA

VB 26	female	58	negRA	18	1.6	4.8	7.5	15	0	--	0.068	0.932	negRA
VB 27	female	57	PSA	1	2.54	6.2	0	0	0	--	0.18	0.82	RA
VB 28	female	55	negRA	12	1.15	1.9	0	0	0	--	0.038	0.962	RA
VB 29	female	40	PSA	16	2.12	10.7	0	20	0	yes	0.177	0.823	RA
VB 30	male	41	PSA	7	1.15	1.9	0	7.5	0	yes	0.609	0.391	PSA
VB 31	male	75	negRA	2	1.84	4.7	2.5	15	0	--	0.603	0.397	PSA
VB 32	female	63	negRA	11	3.1	1.9	0	10	0	yes	0.072	0.928	RA
VB 33	female	52	negRA	0	5.26	1.9	5	10	0	--	0.001	0.999	RA
<b>VB 34</b>	female	63	mixed	0	5.95	78.4	0	0	0	--	0.772	0.228	PSA
VB 35	male	39	PSA	0	1.15	1.9	0	0	0	--	0.889	0.111	PSA