Supplementary methods

S1.1. Analysis of neuroimmune biomarkers

The panel of biomarkers consisted of tumor necrosis factor (TNF), lymphotoxin-alpha (LTα), IL-8, IL-6, IL-1α, IL-1β, and IL-1 receptor antagonist (IL-1Ra) in CSF and plasma and IL-6 receptor (IL-6R), glycoprotein (gp130), TNF receptor 1 (TNFR1), and TNFR2 in CSF and serum. The CRP assay range was 0-10 mg/L, and the limit of detection was 0.02 mg/L. The CVs reported by the manufacturer were 4.1–6.9% intra-assay and 5.8–6.3% inter-assay.
The detection limit of the suPAR assay was 0.4 µg/L. The CV reported by the manufacturer was 2.3–6.0%. The assay range was 1-15 µg/L. Lower limit of detection for NfL and GFAP were 0.065 ng/L and 0.475 ng/L, respectively. The functional lower limit of quantification were 0.8 ng/L for NfL and 16.6 ng/L for GFAP using serum and 8 ng/L for NfL and 166 ng/L for GFAP using CSF. To evaluate and monitor assay performance over time four quality controls two provided by the manufacturer, one in-house prepared serum pool and one in-house prepared CSF pool were included in each run. The total analytical variation for the included controls were; 9-16 % for NfL and 9-18% for GFAP.

S1.2. Sample Processing and 1H NMR data collection

Serum spectra were acquired using a spin-echo sequence (Carr-Purcell-Meiboom- Gill [CPMG]), 32 data collections, an acquisition time of 1.5 s, a relaxation delay of 2 s, and a fixed receiver gain. CSF spectra were acquired with a 1D ^{1}H with Nuclear Overhauser Effect Spectroscopy (NOESY) presaturation scheme (2 s pre-saturation) for attenuation of the water resonance (34). All samples were run within 9 hours of being thawed.

S1.3. Multivariate Statistical Analysis

Models were validated through external cross-validation, whereby the data was separated into train (90%) and test sets (10%), 10 times for every 100 external fold, producing 1000 models in total. The training data was used to learn discriminatory patterns in the predictor

variables, whereas the test set was used for assessing model prediction parameters (accuracy, sensitivity, and specificity). Significant separation between OA and pain-free samples was determined by comparing to a null distribution, where classes were shuffled randomly to achieve the same sized dataset for direct comparison. The key variables for distinguishing between classes were identified using Variable Importance in Projection (VIP) scores and were investigated using univariate analysis as described below.

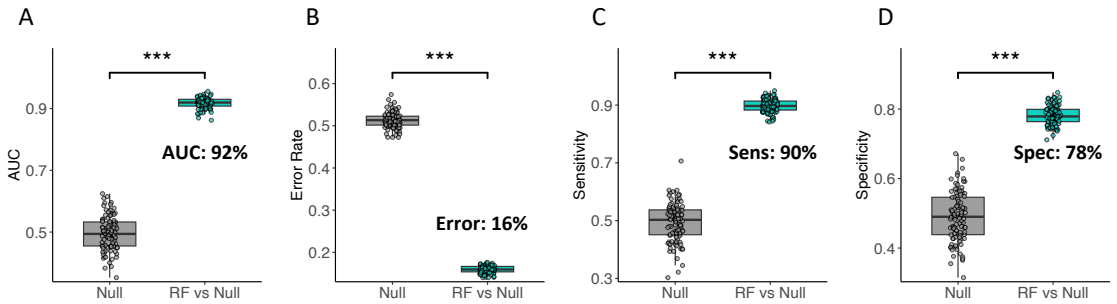
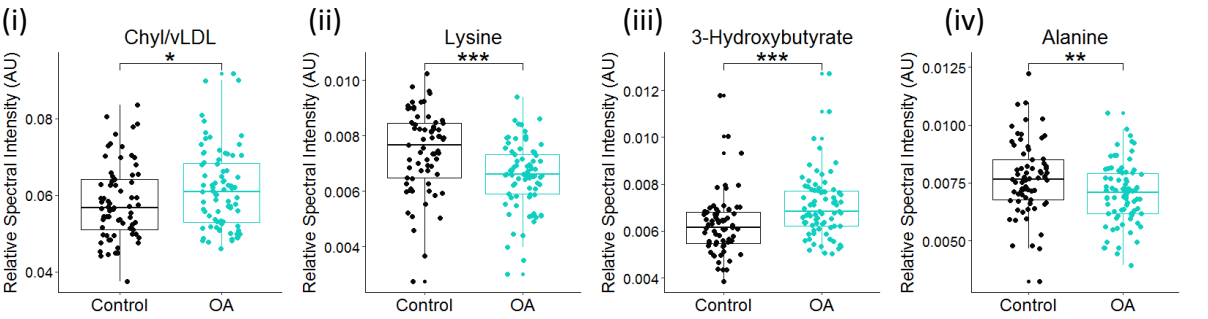


Figure S1. Random forest validation of OA vs control classification. (A) Area under the ROC curve (AUC) for the random forest (RF) model built on the same serum–CSF–protein dataset as Figure 1, compared with a permuted null model. The RF achieved an AUC of 0.92, significantly higher than the null distribution (***p<0.001). (B) Error rate of the RF model was 16%, lower than the null model (***p<0.001). (C) Sensitivity of the RF model was 90%, again exceeding the null model (***p<0.001). (D) Specificity of the RF model was 78%, higher than the null model (***p<0.001). These results show that an orthogonal machine learning approach reproduces the discrimination seen with OPLS-DA, supporting the robustness of the biofluid signature.

A. Serum



B. CSF

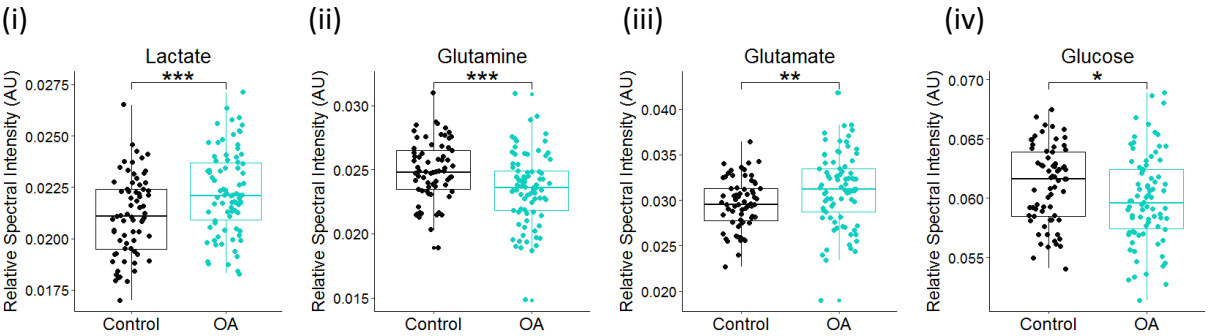
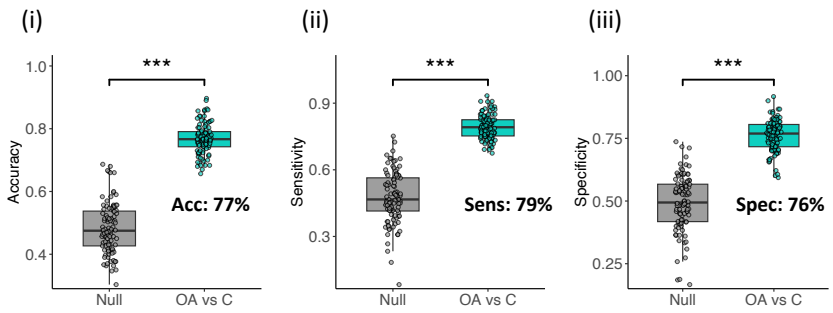


Figure S2. Box plots of top VIP metabolites from the single biofluid OPLS-DA models. (A) Serum. (i–iv) Concentrations of the highest ranking serum metabolites from the OPLS-DA model comparing OA (turquoise) with pain free controls (black): chylomicron/VLDL (i), lysine (ii), 3-hydroxybutyrate (iii) and alanine (iv). All four metabolites contributed strongly to OA–control discrimination in Figure 1A(ii) and remained significantly different on univariate testing. (B) CSF. (i–iv) Corresponding box plots for the top CSF metabolites from the CSF-only model in Figure 1B(ii): lactate (i), glutamine (ii), glutamate (iii) and glucose (iv). CSF lactate and glutamate were higher in OA, while CSF glutamine and glucose were lower. Data are shown as relative spectral intensity (arbitrary units). Asterisks denote significance: *p<0.05, **p<0.01, ***p<0.001 (unpaired t tests).

A



B

VIPs all (A)	VIP score (all)	VIPs Matched cohort (B)	VIP score (matched)
Serum histidine	2.61	**Serum histidine**	2.56
Serum albumin (lysyl moiety)	2.51	**Serum gp130**	2.28
CSF CRP	2.45	**Serum albumin (lysyl moiety)**	2.28
CSF GFAP	2.35	**CSF glucose**	2.25
Plasma IL-6	2.15	**Serum glutamine**	2.20
CSF glucose	2.12	Plasma TNFβ	2.05
Serum glutamine	2.05	**Serum citrate**	1.93
Serum lysine	1.99	CSF glutamine	1.81
Serum GFAP	1.88	**CSF acetate**	1.77
CSF acetate	1.85	Serum Lactate	1.77
CSF creatinine	1.83	CSF IL-8	1.74
CSF glutamate	1.78	**Serum lysine**	1.66
CSF TNFRII	1.70		
CSF supar	1.68		
CSF NfL	1.68		
CSF lactate	1.67		
Serum NfL	1.62		
Plasma CRP	1.53		
Serum 3-hydroxybutyrate	1.52		
Serum citrate	1.51		

Figure S3. Age- and BMI-matched model metrics and comparison of VIP rankings. (A) Performance of the OPLS-DA model built on the age- and BMI-matched subset of osteoarthritis (OA) and control participants. Compared with the permuted null model, the matched model retained good discrimination with (i) accuracy 77%, (ii) sensitivity 79% and (iii) specificity 76% (all ***p<0.001), confirming that group separation is not driven solely by demographic differences. (B) Side-by-side list of variables with the highest variable importance in projection (VIP) scores in the full, unmatched model (left) and in the matched cohort model (centre), together with VIP scores for the matched model (right). Variables that appear in both lists, such as serum histidine, serum albumin (lysyl moiety), CSF glucose, serum glutamine, serum citrate, CSF acetate and serum lysine, represent metabolites and proteins that discriminate OA from controls independently of age and BMI.

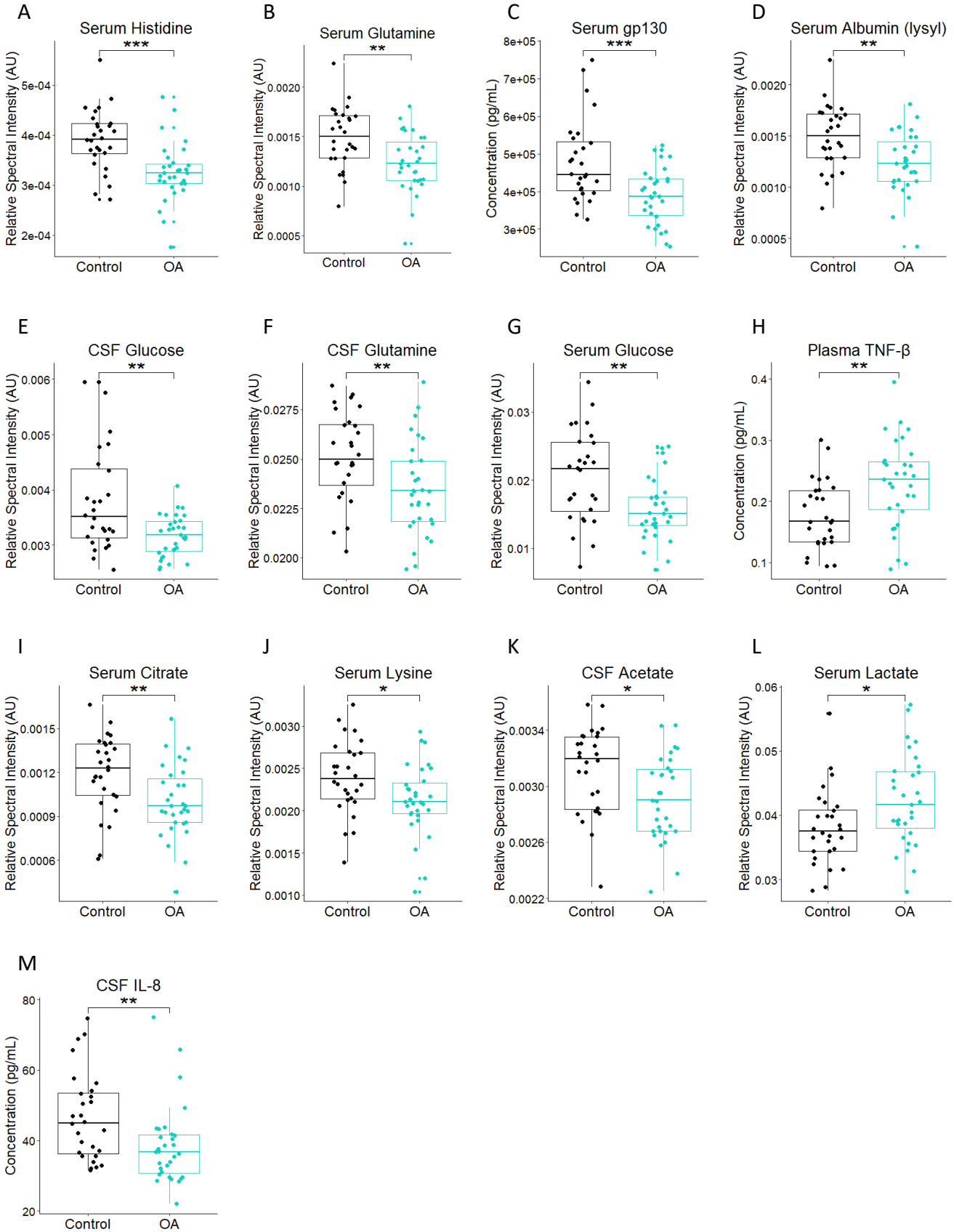


Figure S4. Key serum and CSF metabolites differentiating osteoarthritis and control participants in age- and BMI-matched subgroups. Boxplots display metabolite concentrations in osteoarthritis patients (n=31) and pain-free controls (n=28) matched for age and BMI. Metabolites were selected based on their discriminative value in multivariate analyses from the full cohorts. Asterisks indicate statistical significance based on unpaired t-tests: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$. These findings highlight the robustness of histidine, glutamine, and related metabolite alterations even after accounting for demographic confounders.

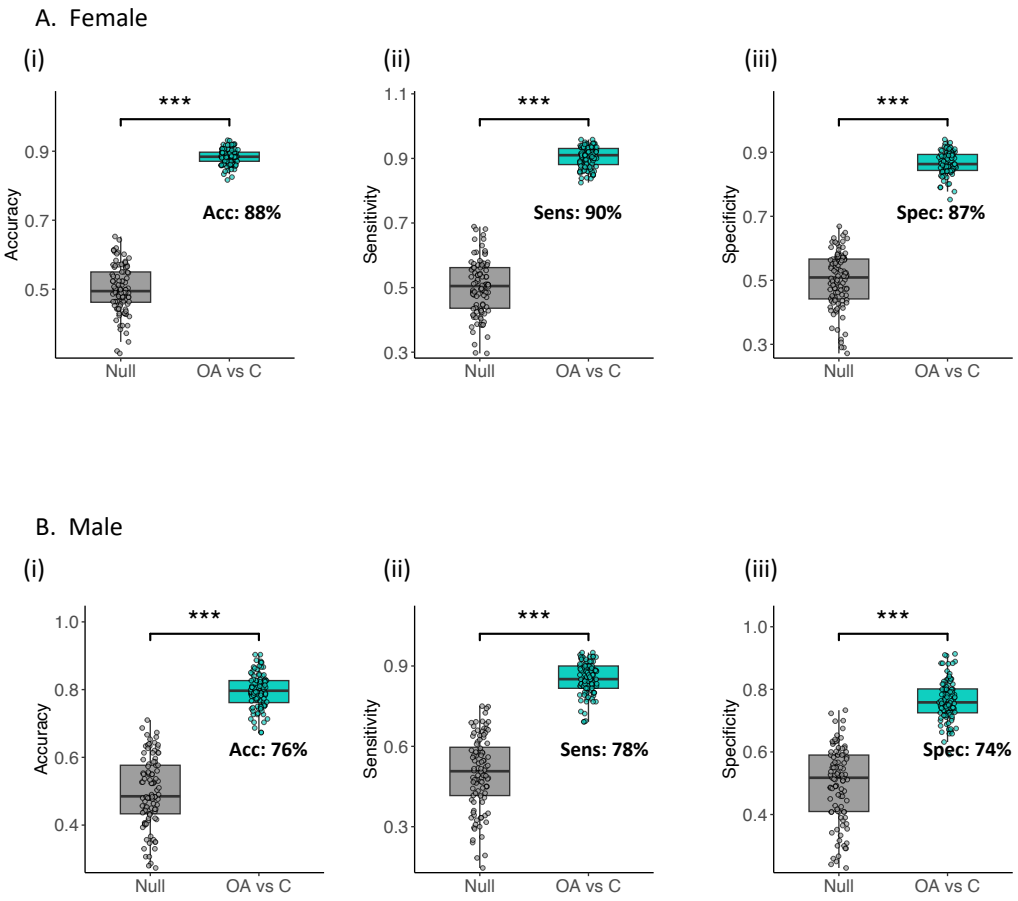


Figure S5. Sex stratified OPLS-DA model performance. (A) Females. Compared with the permuted null model, the female-only OPLS-DA model discriminated osteoarthritis (OA) from controls with (i) accuracy 88%, (ii) sensitivity 90% and (iii) specificity 87% (all ***p<0.001), indicating a strong metabolic–inflammatory signal in women. (B) Males. The male-only model also separated OA from controls, but with lower performance, giving (i) accuracy 76%, (ii) sensitivity 78% and (iii) specificity 74% (all ***p<0.001). These data support the main-figure finding that the OA biofluid signature is more pronounced in females than in males in this cohort.

Variable	Effect	Df	Sum Sq	Mean Sq	F value	p value
Age	Diagnosis	1	33080.1114	33080.11142	203.38	**<0.001**
	Sex	1	549.7755	549.7755	3.38	0.068
	Diagnosis:Sex	1	26.6939	26.6939	0.164	0.686
	Residuals	145	23584.5467	162.65205		
BMI	Diagnosis	1	198.60653	198.60653	8.763	**0.004**
	Sex	1	2.62337	2.62337	0.116	0.734
	Diagnosis:Sex	1	0.12711	0.12711	0.006	0.94
	Residuals	143	3240.90338	22.66366		
SIQR	Diagnosis	1	24792.4501	24792.45005	164.705	**<0.001**
	Sex	1	1366.58583	1366.58583	9.079	**0.003**
	Diagnosis:Sex	1	1361.10329	1361.10329	9.042	**0.003**
	Residuals	140	21073.7497	150.52678		
MDI	Diagnosis	1	284.67421	284.67421	9.239	**0.003**
	Sex	1	40.52171	40.52171	1.315	0.253
	Diagnosis:Sex	1	97.88644	97.88644	3.177	0.077
	Residuals	140	4313.91068	30.81365		
GAD7	Diagnosis	1	1.85585	1.85585	0.252	0.617
	Sex	1	8.3198	8.3198	1.128	0.29
	Diagnosis:Sex	1	10.92888	10.92888	1.482	0.226
	Residuals	140	1032.45103	7.37465		
ISI	Diagnosis	1	649.43103	649.43103	27.98	**<0.001**
	Sex	1	11.97588	11.97588	0.516	0.474
	Diagnosis:Sex	1	60.28657	60.28657	2.597	0.109
	Residuals	140	3249.46624	23.21047		
Serum Albumin	Diagnosis	1	1.52E-05	1.52E-05	32.758	**<0.001**
	Sex	1	2.20E-06	2.20E-06	4.726	**0.031**
	Diagnosis:Sex	1	2.50E-07	2.50E-07	0.541	0.463
	Residuals	144	6.70E-05	4.70E-07		
Serum Glutamine	Diagnosis	1	1.40E-06	1.40E-06	13.959	**<0.001**
	Sex	1	1.20E-07	1.20E-07	1.154	0.285
	Diagnosis:Sex	1	4.60E-07	4.60E-07	4.573	**0.034**
	Residuals	144	1.45E-05	1.00E-07		
Serum Histidine	Diagnosis	1	2.70E-07	2.70E-07	35.529	**<0.001**
	Sex	1	1.00E-08	1.00E-08	0.933	0.336
	Diagnosis:Sex	1	1.00E-08	1.00E-08	1.438	0.232
	Residuals	144	1.08E-06	1.00E-08		
CSF Glutamine	Diagnosis	1	4.48E-05	4.48E-05	6.521	**0.012**
	Sex	1	2.40E-07	2.40E-07	0.035	0.853
	Diagnosis:Sex	1	1.23E-06	1.23E-06	0.179	0.673
	Residuals	145	9.96E-04	6.87E-06		
CSF Glutamate	Diagnosis	1	1.50E-04	1.50E-04	12.776	**<0.001**
	Sex	1	2.00E-05	2.00E-05	1.4	0.239
	Diagnosis:Sex	1	3.00E-05	3.00E-05	2.972	0.087
	Residuals	145	0.00169	1.00E-05		
CSF CRP	Diagnosis	1	0.00812	0.00812	26.473	**<0.001**
	Sex	1	9.00E-05	9.00E-05	0.281	0.597
	Diagnosis:Sex	1	6.00E-05	6.00E-05	0.184	0.669
	Residuals	145	0.04447	3.10E-04		
Plasma CRP	Diagnosis	1	44.68392	44.68392	8.755	**0.004**
	Sex	1	25.22515	25.22515	4.942	**0.028**
	Diagnosis:Sex	1	4.06943	4.06943	0.797	0.373
	Residuals	145	740.07455	5.10396		

Figure S6. Two-way ANOVA of clinical, metabolic and inflammatory variables by diagnosis and sex. Summary table showing the effects of diagnosis (OA vs control), sex, and their interaction (diagnosis × sex) on demographic measures (age, BMI), symptom scores (SIQR, MDI, GAD7, ISI), key serum metabolites (albumin, glutamine, histidine), CSF metabolites (glutamine, glutamate), and inflammatory markers (CSF CRP, plasma CRP). A strong main effect of diagnosis was observed for SIQR, ISI, serum histidine, CSF glutamine, CSF glutamate and CSF CRP (all $p<0.01$), confirming that these variables differ between OA and controls. Sex effects were evident for SIQR and plasma CRP, and significant diagnosis × sex interactions were detected for SIQR and serum glutamine, consistent with the sex specific patterns shown in Figure 5. P values <0.05 are shown in bold.

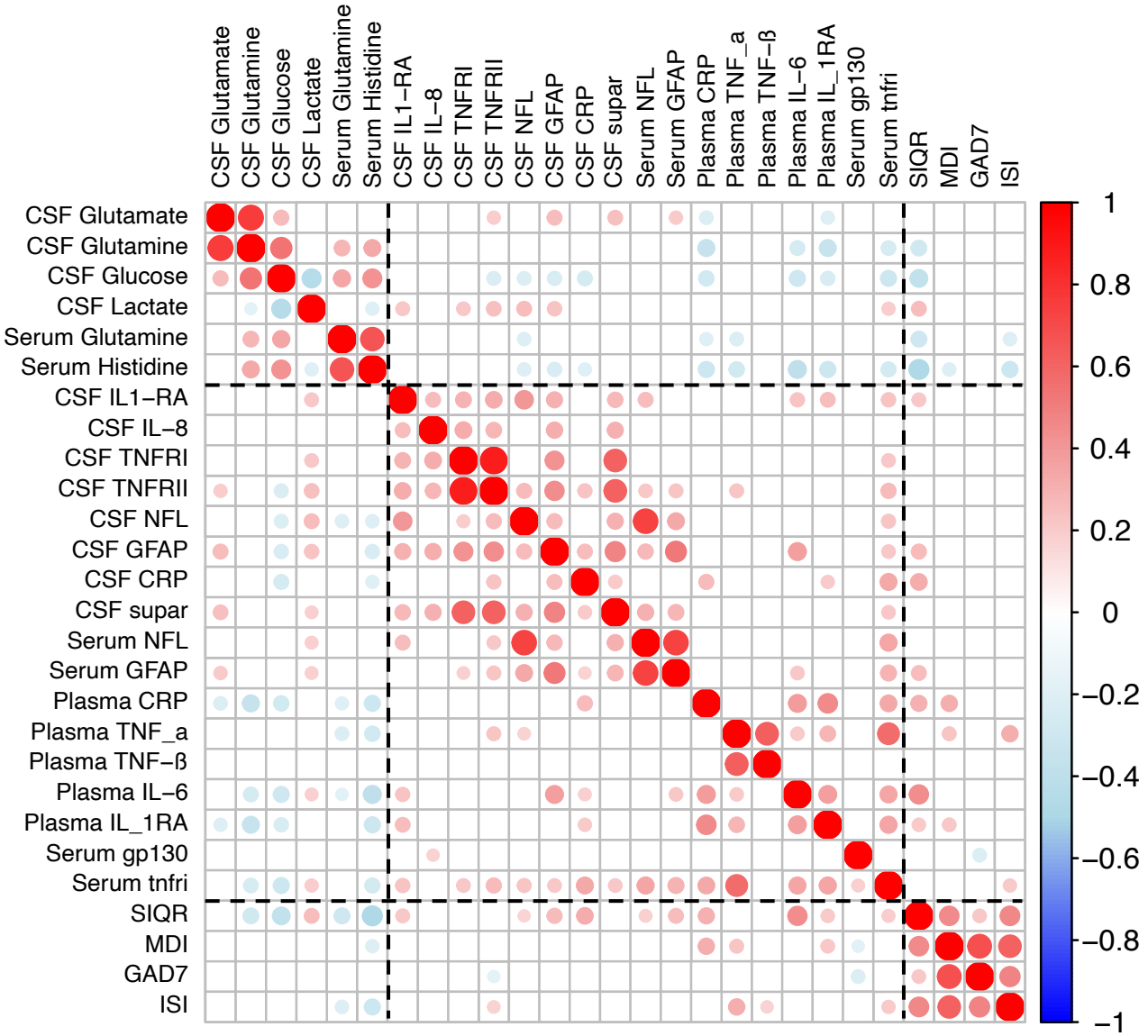


Figure S7. Correlations between key metabolites and protein biomarkers in osteoarthritis and control participants. Heatmap displaying Pearson correlation coefficients between select cerebrospinal fluid (CSF) and serum metabolites and inflammatory or neurodegenerative protein biomarkers across all study participants (OA and controls). Metabolites were selected based on their variable importance in projection (VIP) scores from the OPLS-DA model. Strong correlations were observed between glutamate, glutamine, and lactate with several central and peripheral immune markers, including TNFRI, TNFRII, IL-1RA, and GFAP, suggesting metabolic-immunological interactions relevant to OA pain pathophysiology.