Supplementary Materials

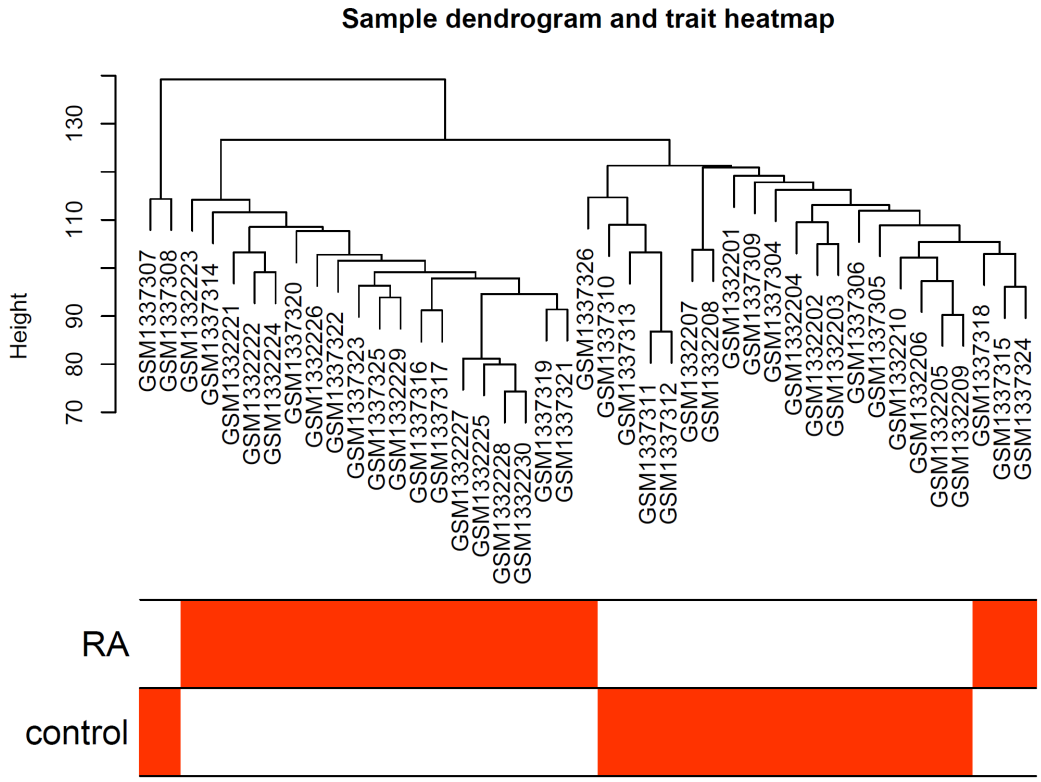

Figure S1. Sample dendrogram and trait heatmap: The upper dendrogram is a hierarchical clustering tree constructed based on the similarity of gene expression profiles among samples, where the length of the branches reflects the differences in expression patterns between samples. The lower color blocks represent the trait heatmap, in which red indicates samples from rheumatoid arthritis (RA) patients and white indicates samples from healthy controls.

Supplementary Methods

WGCNA

Prior to network construction, we implemented rigorous data preprocessing, including normalization using the normalizeBetweenArrays function to minimize technical variation, followed by batch effect correction with the ComBat algorithm from the sva package. We then conducted quality control by removing genes with missing values or insufficient variation (standard deviation = 0), and identified potential sample outliers through hierarchical clustering using the hclust function.

The network construction process began by calculating pairwise gene correlations to generate a similarity matrix, which we transformed into an adjacency matrix using an optimized soft power threshold that satisfies scale-free topology. Through systematic evaluation with the pickSoftThreshold function and visual confirmation of scale-free topology fit (targeting $R^2 = 0.9$), we determined the optimal soft thresholds to be $\beta = 8$ for periodontitis and $\beta = 5$ for RA. The adjacency matrix was subsequently converted to a topological overlap matrix (TOM) to account for shared neighborhood relationships among genes, providing a more biologically meaningful measure of network interconnectedness.

Module detection employed average linkage hierarchical clustering of the TOM-based dissimilarity measure, with module identification performed using dynamic tree cutting. We established stringent parameters for module definition, including a minimum module size of 60 genes and a module eigengene dissimilarity threshold of 0.3 for merging similar modules. The resulting modules were visualized through dendrograms with color-coded assignments using the plotDendroAndColors function. Finally, we conducted comprehensive module-trait relationship analyses to identify disease-associated modules, examining correlations between module eigengenes and clinical traits while evaluating gene significance measures and intramodular connectivity to pinpoint potential hub genes.