Cell Reports Medicine, Volume *5*

Supplemental information

A personalized network framework reveals predictive axis of anti-TNF response across diseases

Shiran Gerassy-Vainberg, Elina Starosvetsky, Renaud Gaujoux, Alexandra Blatt, Naama Maimon, Yuri Gorelik, Sigal Pressman, Ayelet Alpert, Haggai Bar-Yoseph, Tania Dubovik, Benny Perets, Adir Katz, Neta Milman, Meital Segev, Yehuda Chowers, and Shai S. Shen-Orr

Cell Reports Medicine, Volume 4

Supplemental information

A personalized network framework reveals predictive axis of anti-TNF response across diseases

Shiran Gerassy-Vainberg, Elina Starosvetsky, Renaud Gaujoux, Alexandra Blatt, Naama Maimon, Yuri Gorelik, Sigal Pressman, Ayelet Alpert, Haggai Bar-Yoseph, Tania Dubovik, Benny Perets, Adir Katz, Neta Milman, Meital Segev, Yehuda Chowers, and Shai S. Shen-Orr

Figure S1

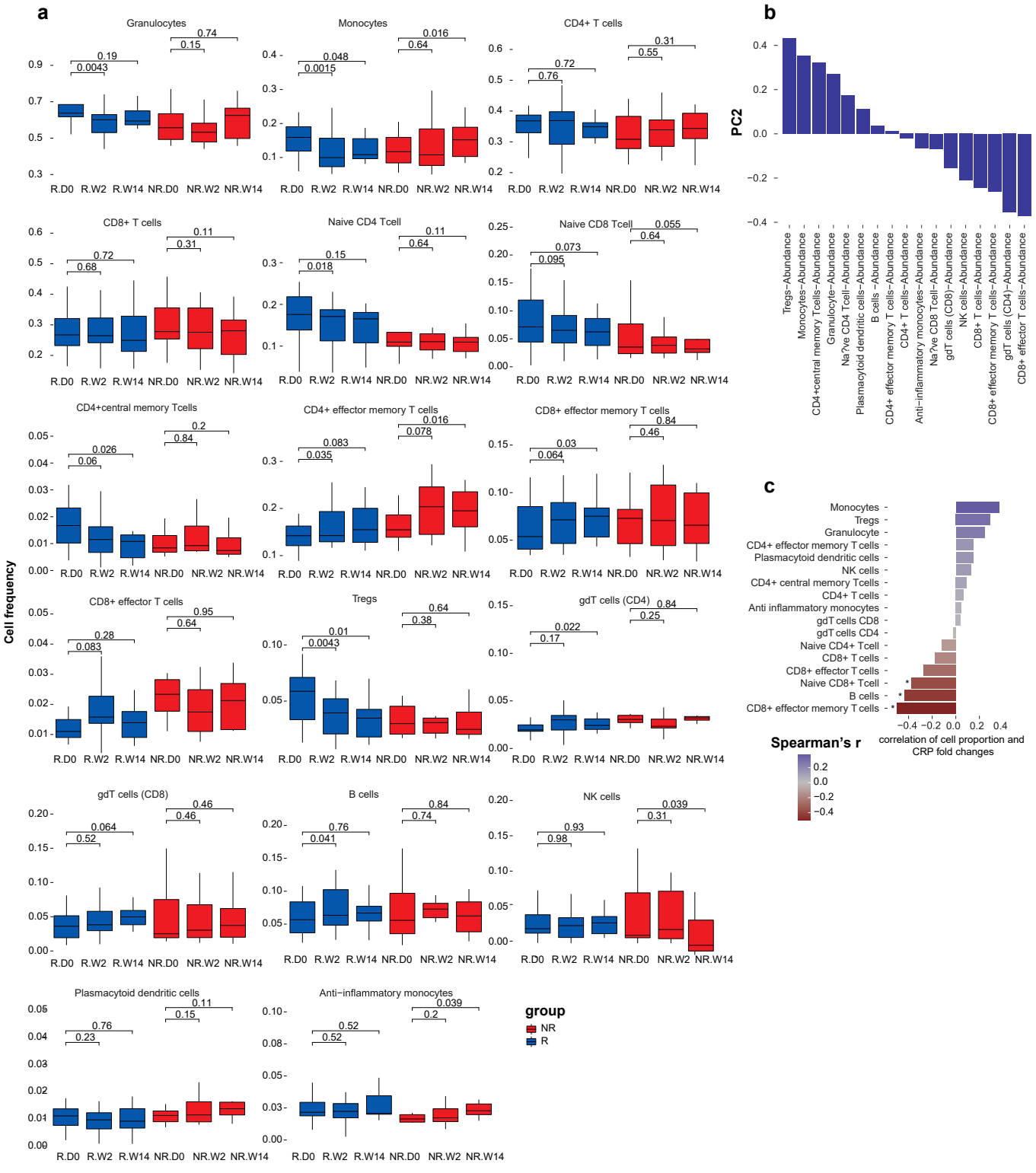

Figure S1. CyTOF reveals multiple cell subset changes in responders following treatment and differences between response groups. Related to Figure 2. a, Cell-type specific alteration in cellular relative abundance during IFX treatment in responders and non-responders (paired-Wilcoxon P-values shown). **b**, Loading plot of PC2 based on major canonical cell composition changes at W2 and W14 compared to baseline. **c**, Correlation of cell abundance changes at W2 and W14 relative to baseline, with changes in CRP (Spearman's correlation coefficients are shown, P-values are calculated by two tailed probability of the t-statistic, P<0.05 for significant p-values).

Figure S2

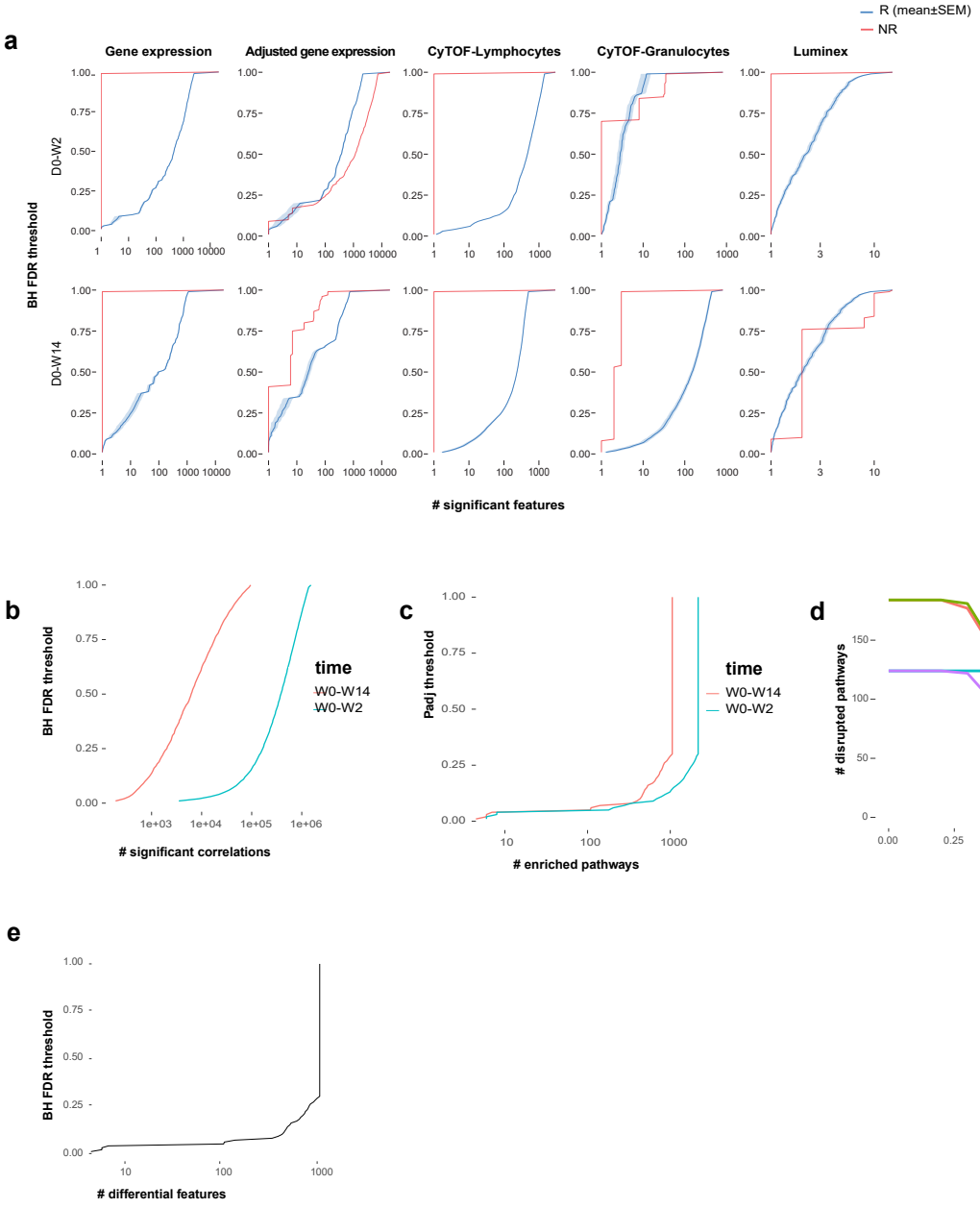

Figure S2. Cumulative FDR plots. **Related to Figures 2-4. a,** The cumulative number of discovered dynamic features, at a range of target FDR values by datatype for each response group**.** Top and bottom panels represent significant changes at W2 and W14 relative to baseline respectively. FDR was calculated using the Benjamini-Hochberg procedure. Responders were subsampled (n=200) to match the non-responder group size. For responders, mean± SEM values are shown. **b**, The cumulative number of significant feature-pair correlations, at a range of FDR values in responders. The correlations were computed using pairwise Spearman's rank correlation, analyzing the D0 -W2 and the D0 -W14 fold-change expression values of the significantly altered features. **c**, The cumulative number of significant enriched pathways, at a range of padj values in responders. The gene set enrichment analysis was performed using fGSEA (nperm=1000, minSize=10, maxSize=400). **d**, The cumulative number of significantly disrupted pathways, at a range of percentile values of the different disruption measures. Significantly disrupted pathways are presented as a function of a percentile cutoff in the different disruption parameters. Purple line indicates the number of disrupted pathways with agreement across the three disruption parameters. **e**, Cumulative plot of the number of differentially expressed genes at baseline per FDR threshold. The genes of the disrupted dynamic pathways were included in the analysis.

Figure S3

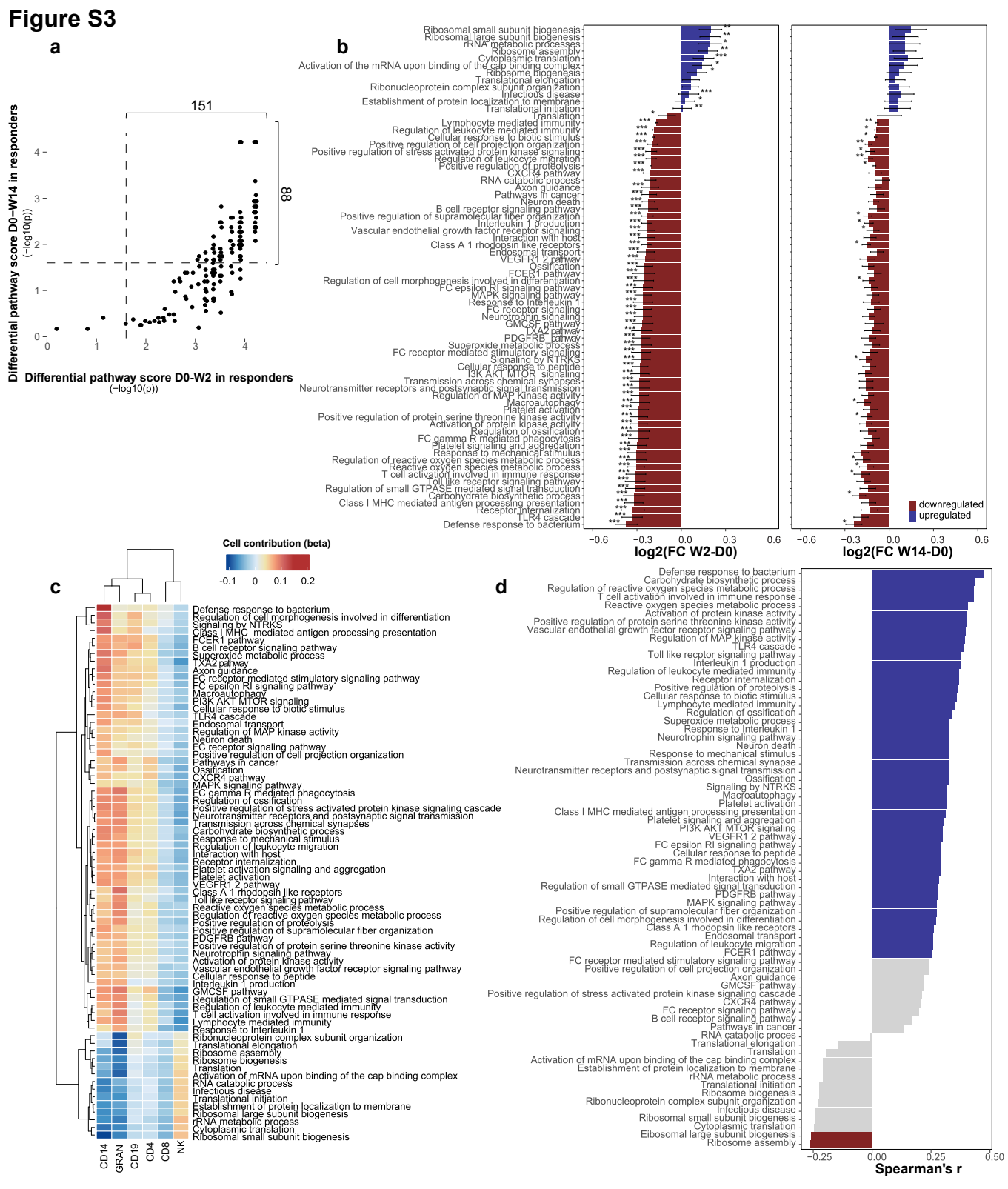

Figure S3. Functional pathways associated with IFX response. Related to Figure 2. a, Scatterplot of p-values obtained by a comparison of pathway scores between W2 and baseline against those obtained by comparing W14 to baseline (-log10 of paired-Wilcoxon P-values shown). Only globally enriched and network connected pathways were included. **b**, Pathway score related dynamics between W2 and W14 relative to baseline. Top 70 pathways are shown. Pathways are ordered by fold change effect size. P-values for pathway score differences between time points were calculated by paired-Wilcoxon test. Significance was determined by FDR<0.05 (Benjamini-Hochberg procedure). **c**, Heatmap representing a cell-specific contribution for the change in the dynamic pathways. The contribution was determined for each gene in the pathway by regressing its unadjusted fold change expression over the major peripheral cell type frequencies. The reported values represent the mean of the coefficients across all genes in the pathway. **d**, Correlation of pathway score expression with CRP. All time point and response groups are included. (Spearman's correlation coefficients are shown, P-values are calculated by two tailed probability of the t-statistic, Pathway which significantly correlated with CRP (FDR<0.05, Benjamini-Hochberg procedure) are colored).

Figure S4

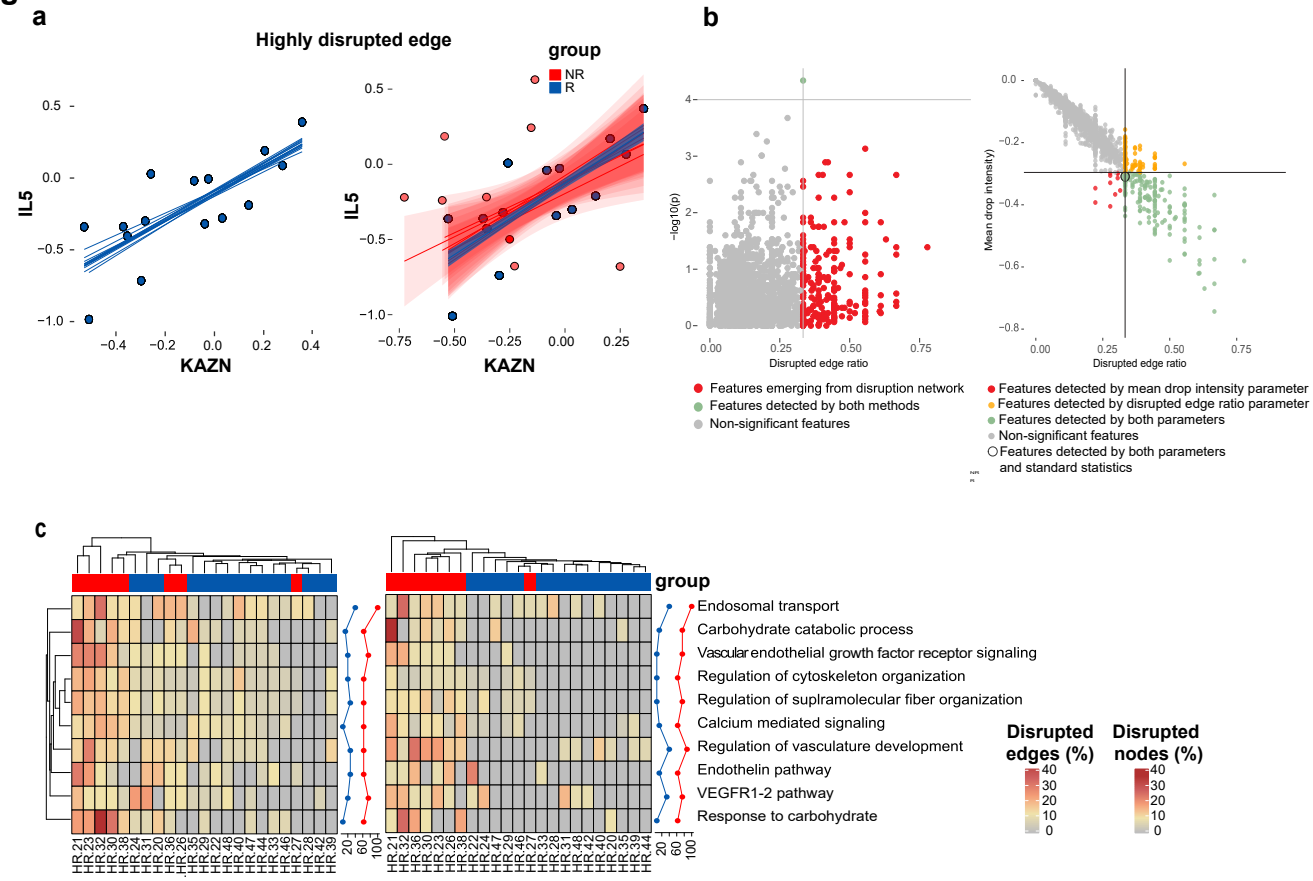

Figure S4. Additional disruption parameters and comparison of the differential signal between response groups dynamics as obtained by the 'Disruption Networks' framework and standard statistics. Related to Figure 3. a, Representative highly disrupted edge demonstrating significant dropout values for non-responders. **b,** Feature-specific differential signal between responders and non-responders' dynamics at W2 relative to baseline, based on the disruption parameters and standard statistics. Left panel, top disrupted edge ratio (x axis, FDR<0.1 for dropout significance and 10th top percentile of disrupted edge ratio) and standard statistics by Wilcoxon test (y axis, FDR<0.1); Right panel, Scatterplot showing feature specific disruption parameters of mean drop intensity against disrupted edge ratio. Points are colored by quartile thresholds (FDR<0.1 for dropout significance and 10th top percentile of the specific disruption parameter). The feature which agreed with the disruption parameters and standard Wilcoxon test is marked with black border. **c,** Aggregation of 'Disruption Networks' statistic across pathways to estimate sample specific disruption in the functional level, according to percentage of disrupted edges and (left) percentage of disrupted nodes (right). Heatmaps represent the disrupted dynamics in each parameter for each pathway and sample at W2 compared to baseline. Top significantly disrupted pathways are presented, defined as those with a complete agreement of all three parameters in the 0.8 percentile. Line graphs describe the percentage of disrupted patients in each response group.

Figure S5

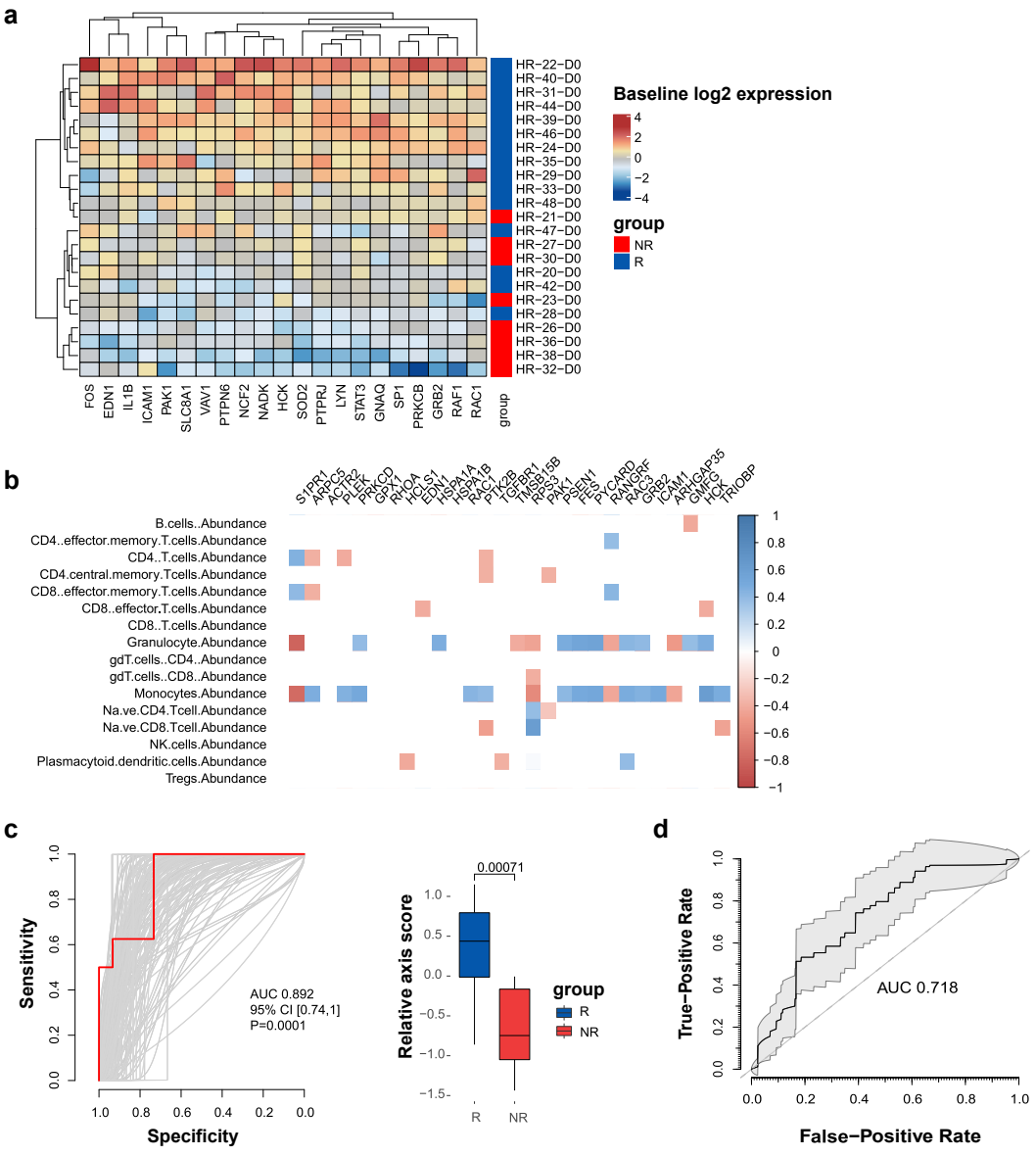

Figure S5. Baseline differences of the significantly dynamics disrupted pathways. **Related to Figures 4-5. a**, Heatmap representing the feature-level baseline differences among genes in the dynamics meta-disrupted pathway (FDR<0.1, Wilcoxon test). **b**, Correlation between the canonical cellular frequencies as obtained by CyTOF, and the bulk unadjusted expression of the fiber organization related genes in responders (Spearman's correlation coefficients are shown, P-values are calculated by two tailed probability of the t-statistic). Only significant correlation values are shown (P<0.05 and |r|≥0.5). **c,** Baseline prediction of IFX response in the primary IFX cohort based on the expanded fiber organization predictive signature score, in the cell adjusted space. Left panel, receiver operating characteristic (ROC) plots of 200-bootsraps. The predictive signature was determined using elastic net (a=0.5, lambda=0.26, 100 repeated 2-fold CV) based on the adjusted baseline differential fiber organization related genes. Significance was determined by permutation test (n perm=10000). Right panel, boxplots of the fiber organization predictive signature score pre-treatment, in the different response groups in the cell-centered bulk expression. **d,** Meta-ROC presenting the predictive performance of three independent public RA cohorts after subsampling to balance the response groups, by sampling n=10 from each response group, 100 times, for each of the RA cohorts.

Figure S6

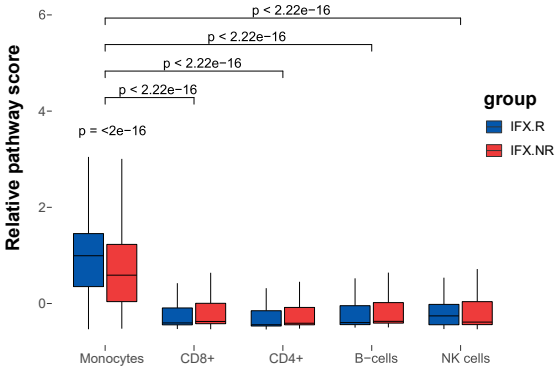

Figure S6. scRNA-seq based comparison of the baseline fiber organization related expression between the main cell-types and response groups. Related to Figures 4. The fiber organization scaled score based on its baseline differential genes was compared between PBMCs major cell types and between response groups (Wilcoxon P-values shown).

Figure S7

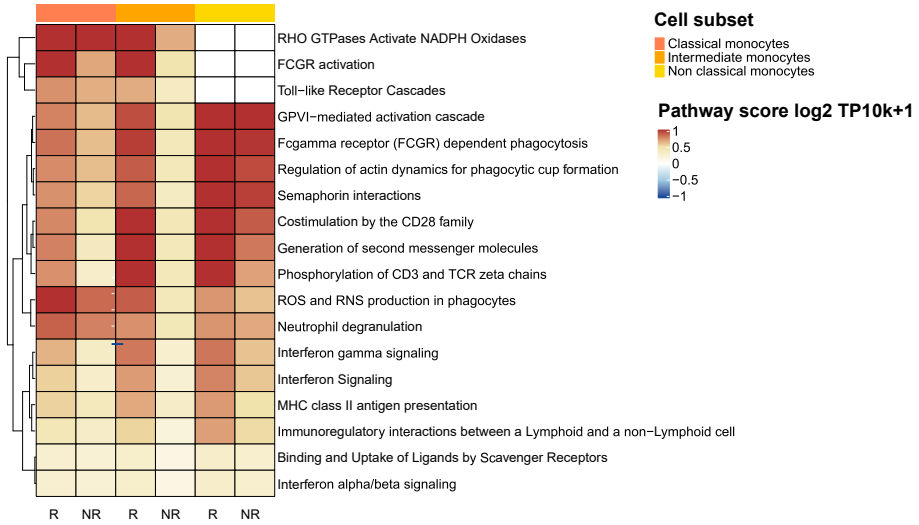

Figure S7. Intermediate monocytes functional pathways associated with the predictive fiber organization signature. Related to Figures 4-5. Heatmap representing the top 20 intermediate-monocytes specific enriched pathways associated with the predictive fiber-organization related signature is shown. Pathways were determined by co-expression network based on the pre-treatment expression of the signature predictive genes in intermediate monocyte based on the scRNA-seq data in each response group followed by enrichment analysis (Spearman's correlation, thinning net by 0.1 top percentile, P-adjust<0.05 for functional enrichment significance by hypergeometric test). Pathways displaying significant differences between response groups in each cell subset are colored (FDR<0.05 by Wilcoxon test).

Figure S8

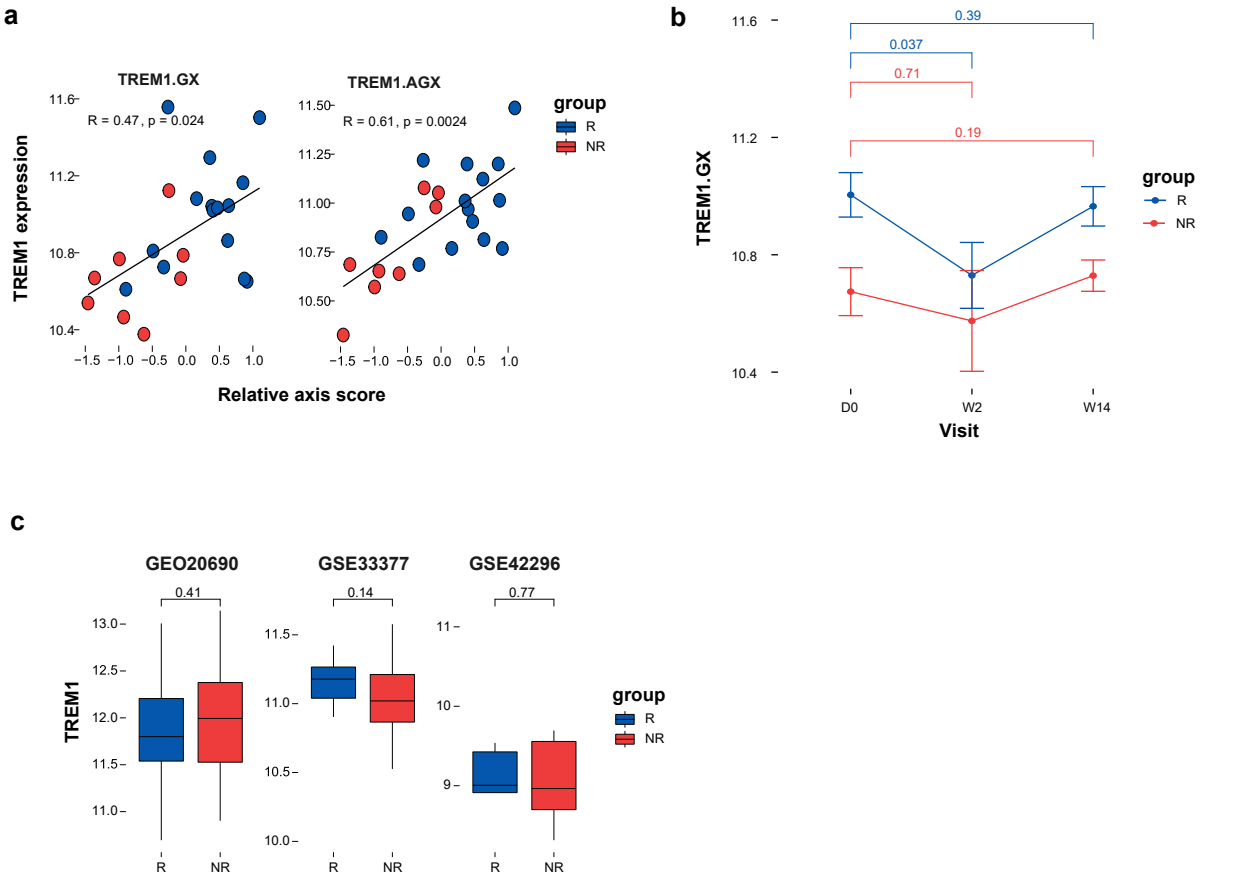

Figure S8. Comparison of the fiber organization predictive signature to previously reported anti-TNF biomarker TREM1. **Related to Figure 5. a,** Scatterplots presenting the correlation between the identified fiber organization predictive score and the expression levels of TREM1, in the primary CD cohort from blood, at baseline. Bulk (GX) and cell centered (AGX) expression are presented at the right and left panels correspondingly. Spearman's correlation coefficients are shown; P-values are calculated by a two-tailed test. **b,** Bulk expression of TREM1 in blood across visits during infliximab treatment in responding and non-responding patients (Wilcoxon one-tailed P-values shown). **c,** Boxplots comparing bulk TREM1 expression between responding and non-responding patients in three independent public RA cohorts (Wilcoxon one-tailed P-values shown).

Figure S9

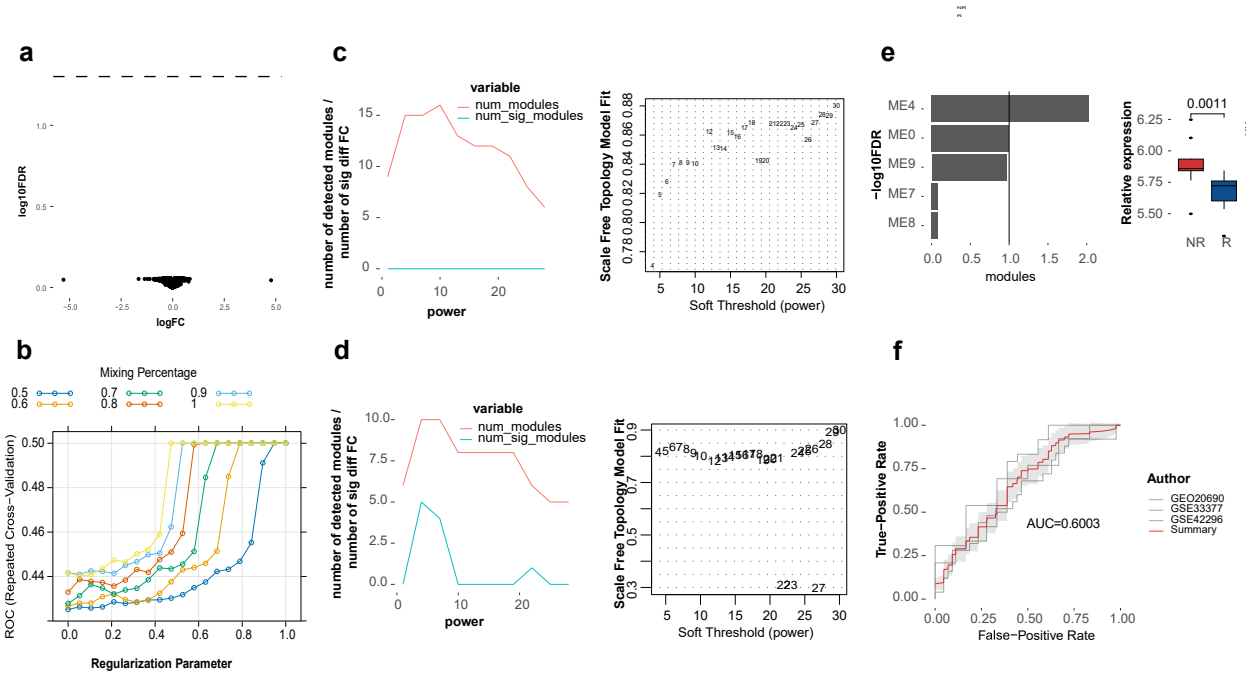

Figure S9. Comparison of Disruption Networks predictive performance to existing methods. **Related to Figure 3. a,** Comparison to limma based differential expression. Volcano plot of differentially expressed genes between responders and non-responders at baseline. Differences are computed post adjustment to cell frequency using the limma R package. Vertical line indicates 0.05 FDR threshold. **b,** Elastic-net ROC results in different parameters of alpha and lambda. Elastic net was computed with 100` repeats of 2-fold cross validation. **c,** comparison to WGCNA. left panel, responders' fold change co-expression network and module identification using WGCNA. Number of observed modules and differential modules between responders and non-responders are shown. Right panel, the corresponding scale free topology model fit per soft threshold. **d,** Left panel, number of modules and differential modules in WGCNA based co-expression network using both responders and non-responders samples. Right panel, as in c. **e,** Baseline differential modules between the response groups, and the differential expression in the gene level. **f,** Meta-ROC presenting the predictive performance of three independent public RA cohorts based on the WGCNA signature.