

Supplemental Material

Multi-cell type gene co-expression network analysis reveals coordinated interferon response and cross cell-type correlations in systemic lupus erythematosus

Bharat Panwar^{1*}, Benjamin J. Schmiedel¹, Shu Liang¹, Brandie White¹, Enrique Rodriguez³, Kenneth Kalunian², Andrew J. McKnight³, Rachel Soloff³, Gregory Seumois¹, Pandurangan Vijayanand^{1,2*}, Ferhat Ay^{1,2*}

¹La Jolla Institute for Immunology, La Jolla, California, USA.

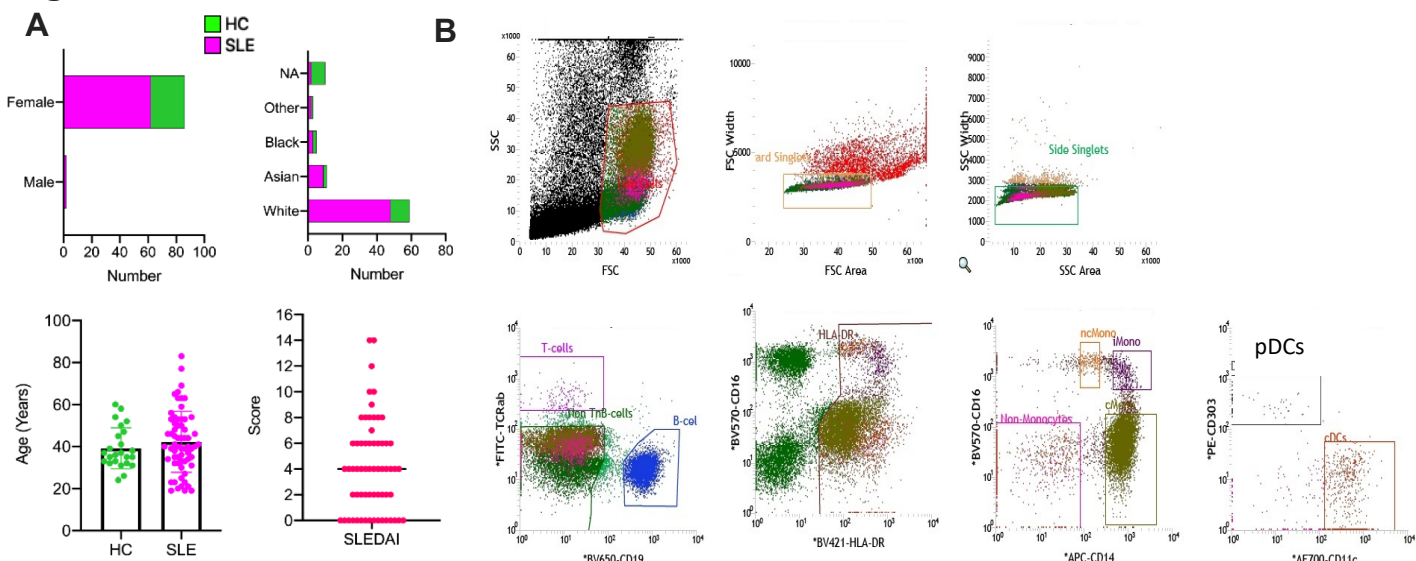
²School of Medicine, University of California San Diego, La Jolla, California, USA.

³Kyowa Kirin Pharmaceutical Research, Inc., La Jolla, California, USA.

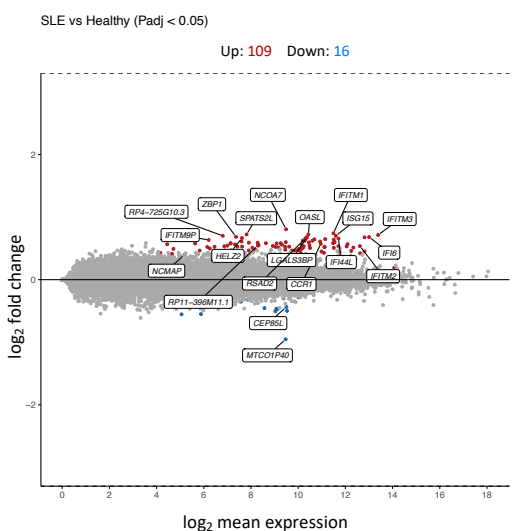
*co-corresponding authors: spanwar@lji.org, vijay@lji.org, ferhatay@lji.org

SUPPLEMENTAL FIGURES

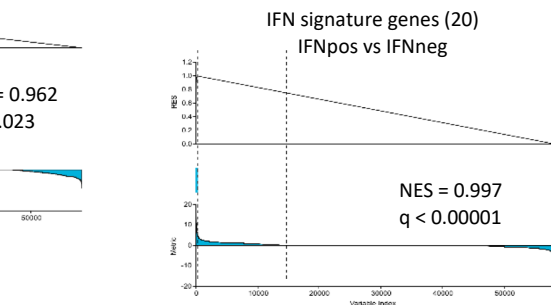
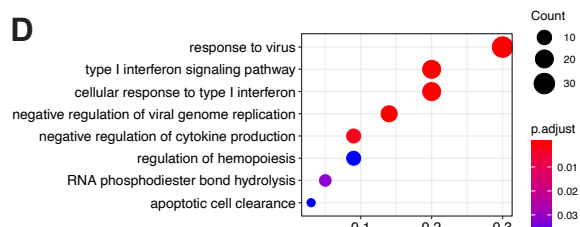
Figure S1



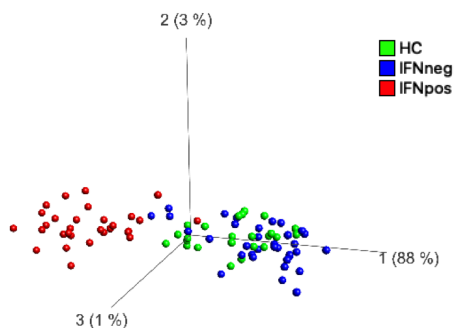
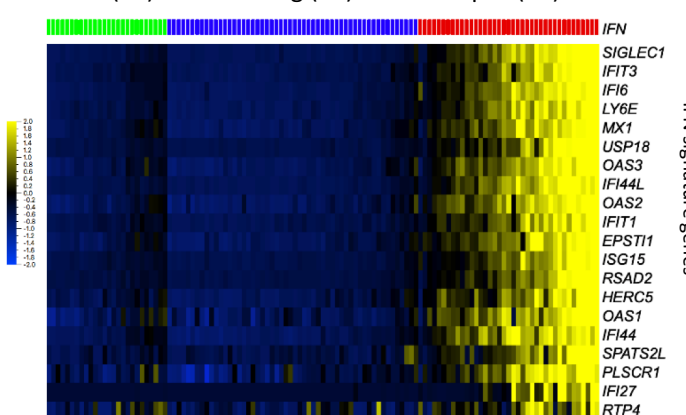
C



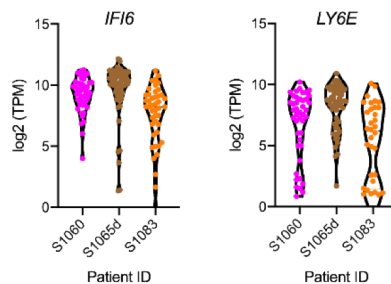
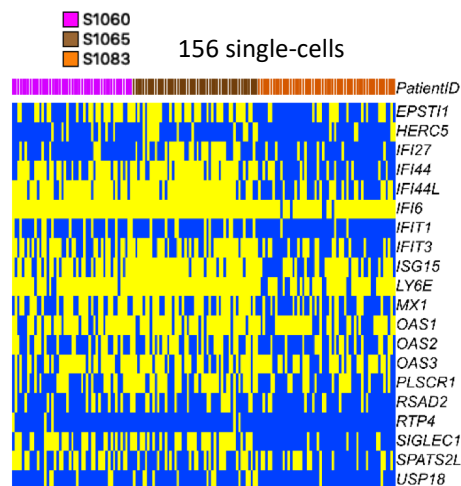
D



E



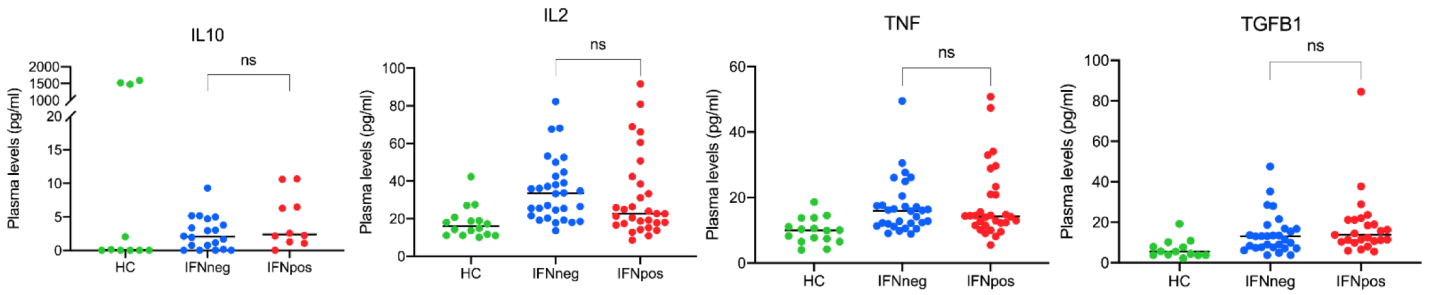
F



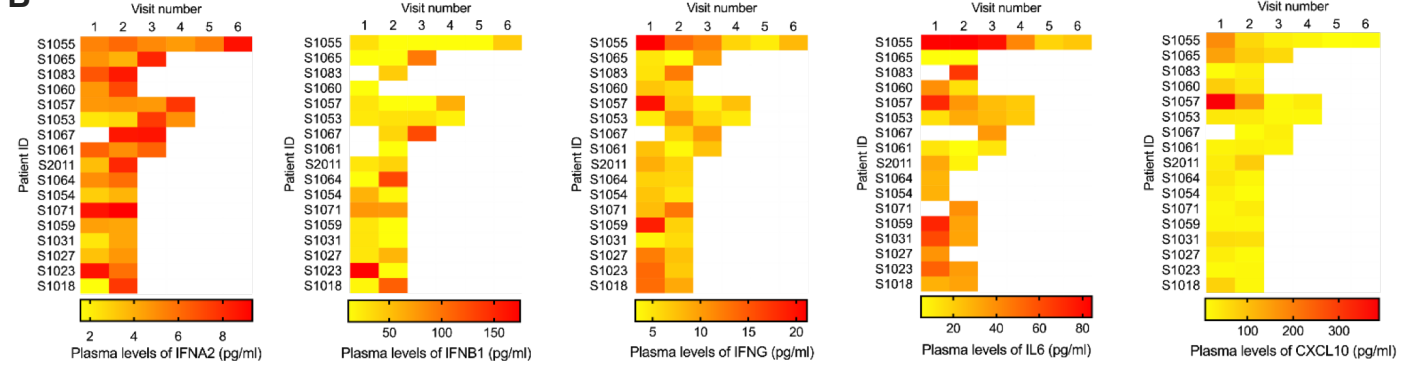
Supplemental Figure S1. Transcription profile of classical monocytes reveals two molecular subtypes of SLE that shows stability in longitudinal data. **(A)** Different plots show the distribution of gender, ethnicity, age, and SLEDAI score in the SLE cohort. Healthy control (HC) samples are highlighted in green color and SLE samples are highlighted in magenta color (details in Supplemental Table S1). **(B)** Gating scheme of T-depleted PBMC for sorting of multiple immune cell types. Live cells were gated based on their forward side scatter versus side scatter profile followed by single cells gating by FSC then SSC and then finally according to their cell surface phenotype as follows, B cells (CD19⁺, TCRa/b⁻), classical Mo (cMono, TCRa/b⁻, CD19⁻, HLA-DR⁺, CD14⁺, CD16⁻), intermediate Mo (iMono, TCRa/b⁻, CD19⁻, HLA-DR⁺, CD16⁺, CD14⁺), non-classical Mo (ncMono, TCRa/b⁻, CD19⁻, HLA-DR⁺, CD16⁺ CD14^{low}), plasmacytoid dendritic cells (pDC, TCRa/b⁻, CD19⁻, HLA-DR⁺, CD303⁺, CD11c⁻) and conventional dendritic cells (cDC, TCRa/b⁻, CD19⁻, HLA-DR⁺, CD303⁻, CD11c⁺). **(C)** The MAplot shows 125 DEGs (P.adj < 0.05 from Benjamini-Hochberg test in DESeq2) between SLE and HC in classical monocytes. The top 20 genes (based on P.adj) are highlighted in boxes. The GSEA plot shows significant enrichment (NES=0.962; q=0.023) of 20 IFN-signature genes (IFN-20) in SLE vs HC. **(D)** Functional annotations (generated by clusterProfiler) of DEGs between IFNpos-vs-IFNneg (top). The color shows the significance (in terms of P.adj), the size is gene counts in annotation, and the X-axis shows gene ratio. The GSEA plot shows significant enrichment (NES=0.997; q<0.00001) of 20 IFN-signature genes (IFN-20) in IFNpos vs IFNneg comparison. **(E)** Expression of IFN-20 genes in all SLE patients (including longitudinal visits). Where each gene is presented as row-wise z-scores of transcripts per million (TPM) in IFNpos (red), IFNneg (blue) and HC (green); each column represents an individual patient and patients were clustered within their IFN status based on PC1 values (top). The PCA plot (bottom) shows the two molecular sub-types of SLE in different colors. Green, blue, and red represent HC, IFNneg and IFNpos, respectively (we used this color scheme in all figures). **(F)** The binary heatmap plot of single cell level expression of IFN-20 in cMo (156 cells) from three different IFNpos patients (shown in different colors in columns). In this binary heatmap, expression level of >1 TPM is highlighted in yellow and <1 TPM is shown in blue. The expression of two genes (*IFI6* and *LY6E*) is also provided as violin plots.

Figure S2

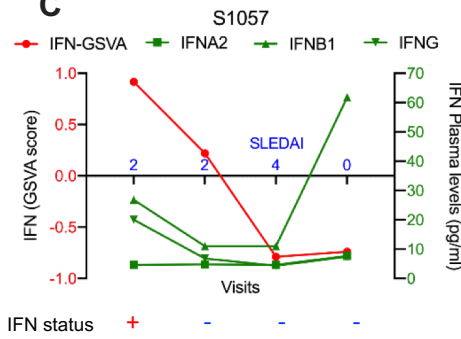
A



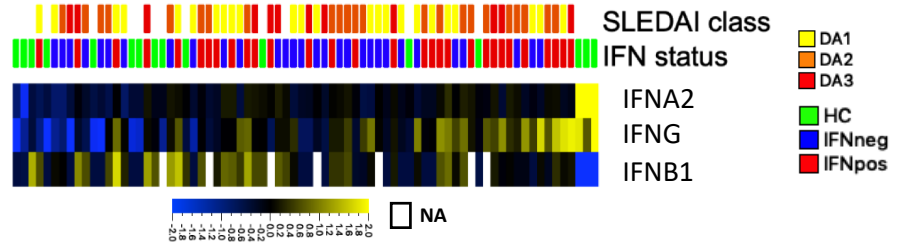
B



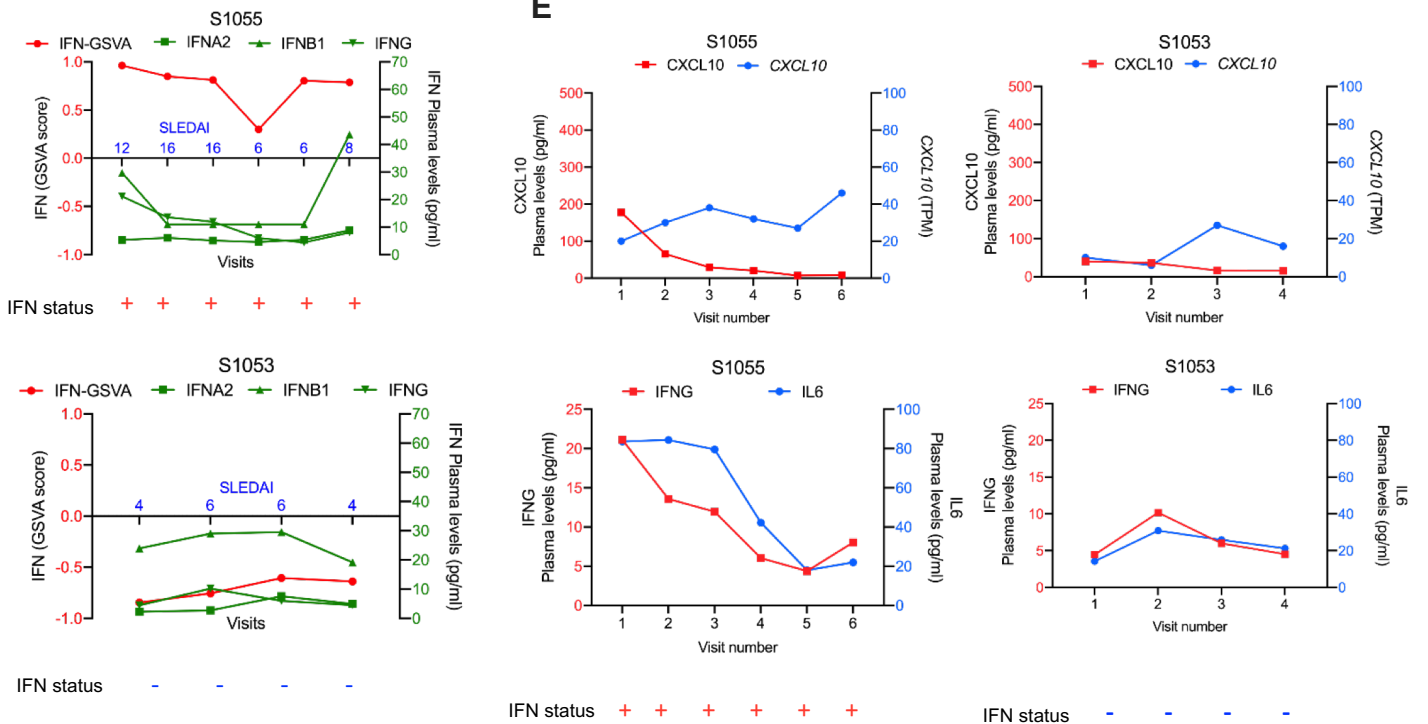
C



D

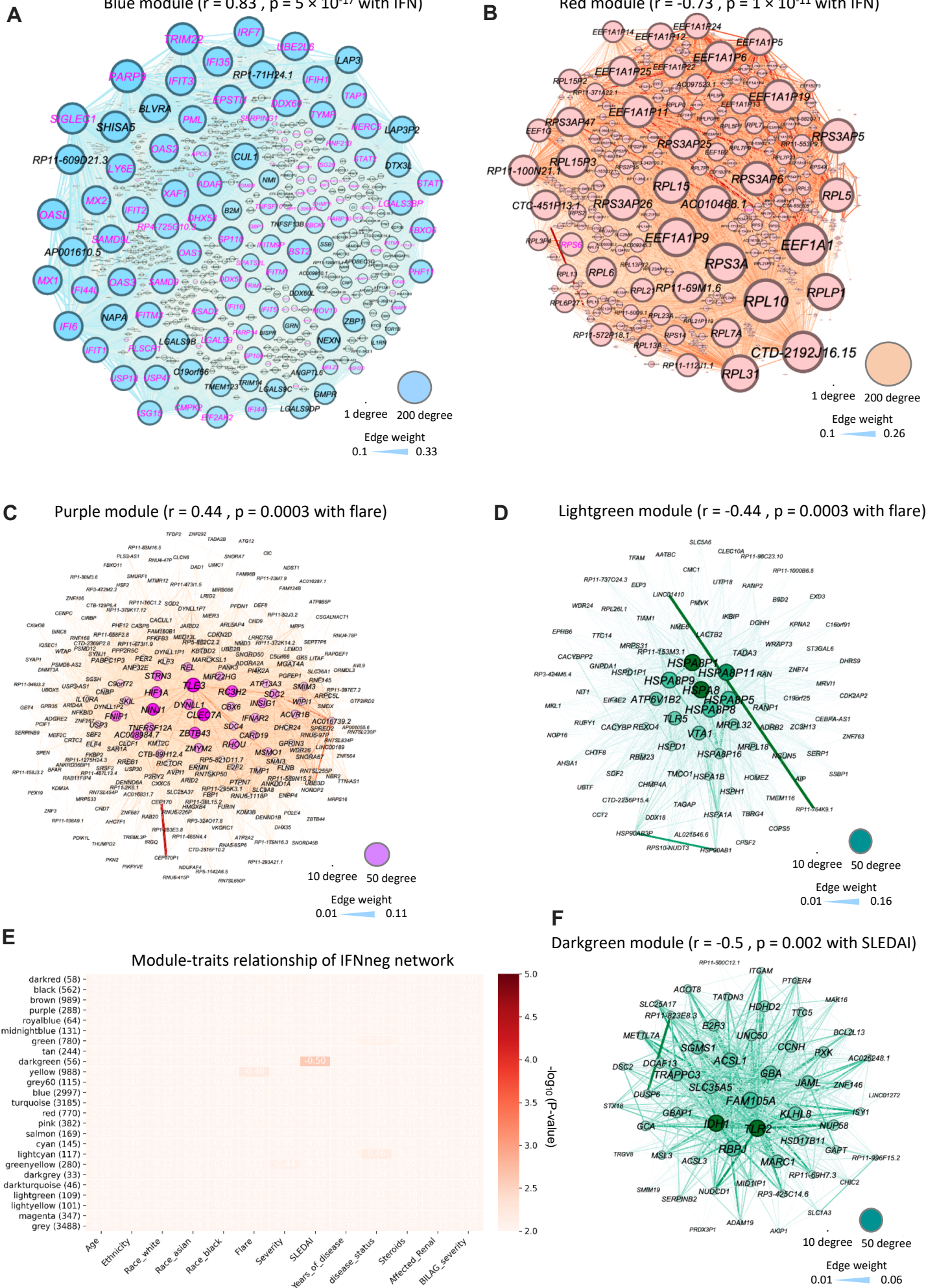


E



Supplemental Figure S2. Expression levels of different IFNs and cytokines in first and longitudinal visits. (A) Scatter dot-plots show plasma levels (pg/ml) of individual cytokines in HC, IFNneg and IFNpos SLE patients. Only first visit samples (n=76) were used in these scatter dot-plots. Differences between IFNpos and IFNneg were calculated using unpaired *t*-test (two-tailed) and statistical significance (p-value) levels are shown in each plot (ns: not significant, *: <0.05). **(B)** Series of heatmaps showing longitudinal information of different IFNs and pro-inflammatory cytokines (IL6 and CXCL10) in plasma. **(C)** Correlation between classical monocytes based IFN feature (GSVA score) and different IFN levels in plasma for three patients (S1057, S1055, and S1053). The SLEDAI score is also provided for each visit. **(D)** Heatmap of different IFN levels in plasma of SLE patients and healthy control. Where each IFN is presented as row-wise z-scores of IFN level in plasma (pg/ml) in IFNpos (red), IFNneg (blue) and HC (green); each column represents an individual patient. All patients are clustered based on PC1 values. The SLEDAI class panel shows SLEDAI score and is divided into three different categories based on score: DA1 (0-2), DA2 (3-7), and DA3 (>7) (yellow, orange and red, respectively). **(E)** Plots with connecting lines (upper and lower) show expression changes of multiple cytokines and chemokines in the longitudinal data for patient S1055 and S1053, where different analytes are plotted on different Y-axes (IFNG on Y1-axis in red color and IL6 on Y2-axis in blue color) in different colors (bottom graphs). The RNA-level expression of *CXCL10* is also shown (Y2-axis) with protein-level expression of *CXCL10* (Y1-axis) (top graphs). The lower panel of bottom plots show their IFN response status over multiple longitudinal visits.

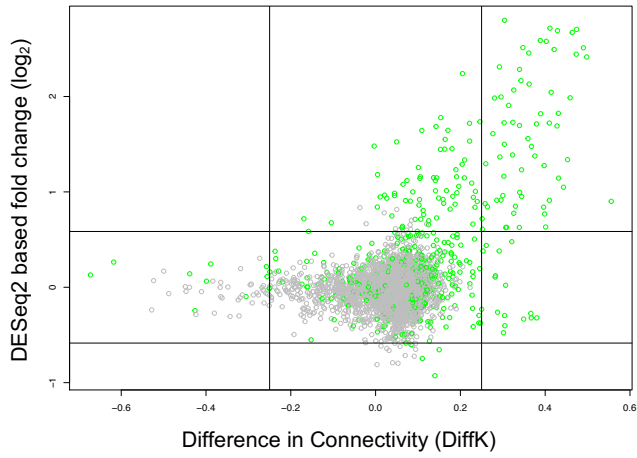
Figure S3



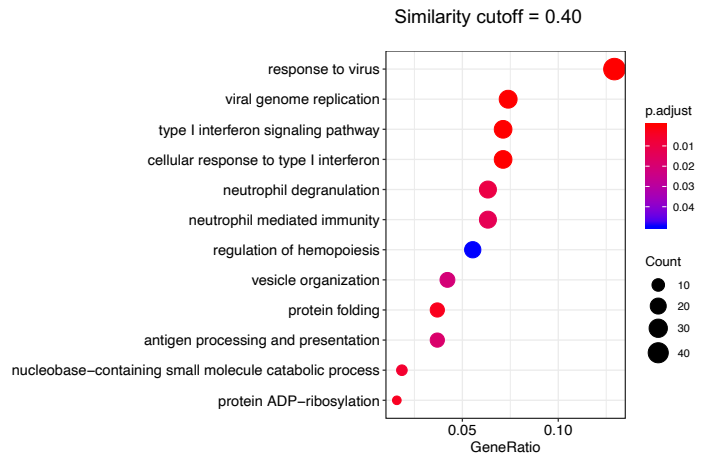
Supplemental Figure S3. Interesting modules from SLE and IFNneg patients based WGCNA network. Gephi based visualization of the blue module **(A)** and red module **(B)**, where nodes are sized according to the number of edges (connections), and the edge thickness is proportional to the strength of co-expression. Available IFN related genes (IFN-363) are highlighted in magenta colors. Similarly, purple **(C)** and lightgreen **(D)** modules from SLE network are also visualized by Gephi. **(E)** The module-traits correlation matrix heatmap of WGCNA based on total 25 different modules and external clinical features such as age, ethnicity, flare, severity etc from IFNneg patients. The color shows significance ($-\text{Log}_{10}$ of p-value) and Spearman's correlation values are provided in the heatmap. We only show the p-value for modules with significant correlation (p-value < 0.01). The number of genes in each module is also shown (in brackets) with the module name. **(F)** Gephi based visualization of the darkgreen module from IFNneg network, where nodes are sized according to the number of edges (connections), and the edge thickness is proportional to the strength of co-expression.

Figure S4

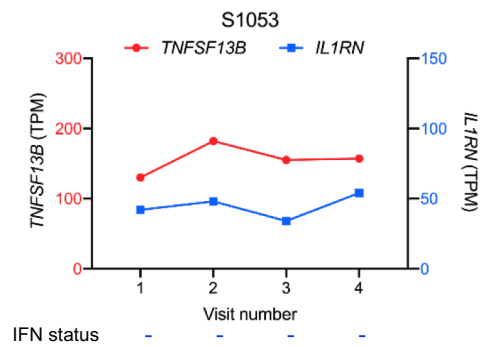
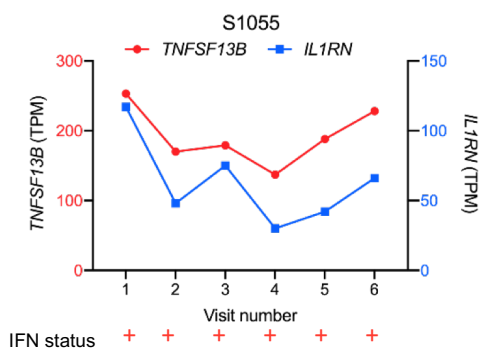
A Differential Network Analysis of IFNpos vs IFNneg (color by IFNpos network)



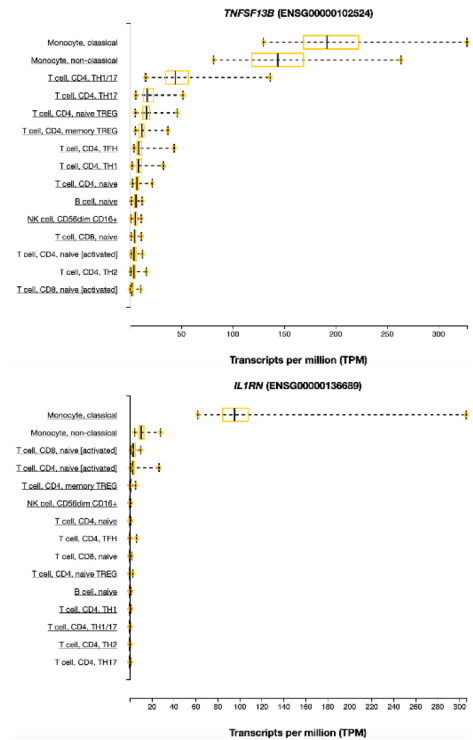
B



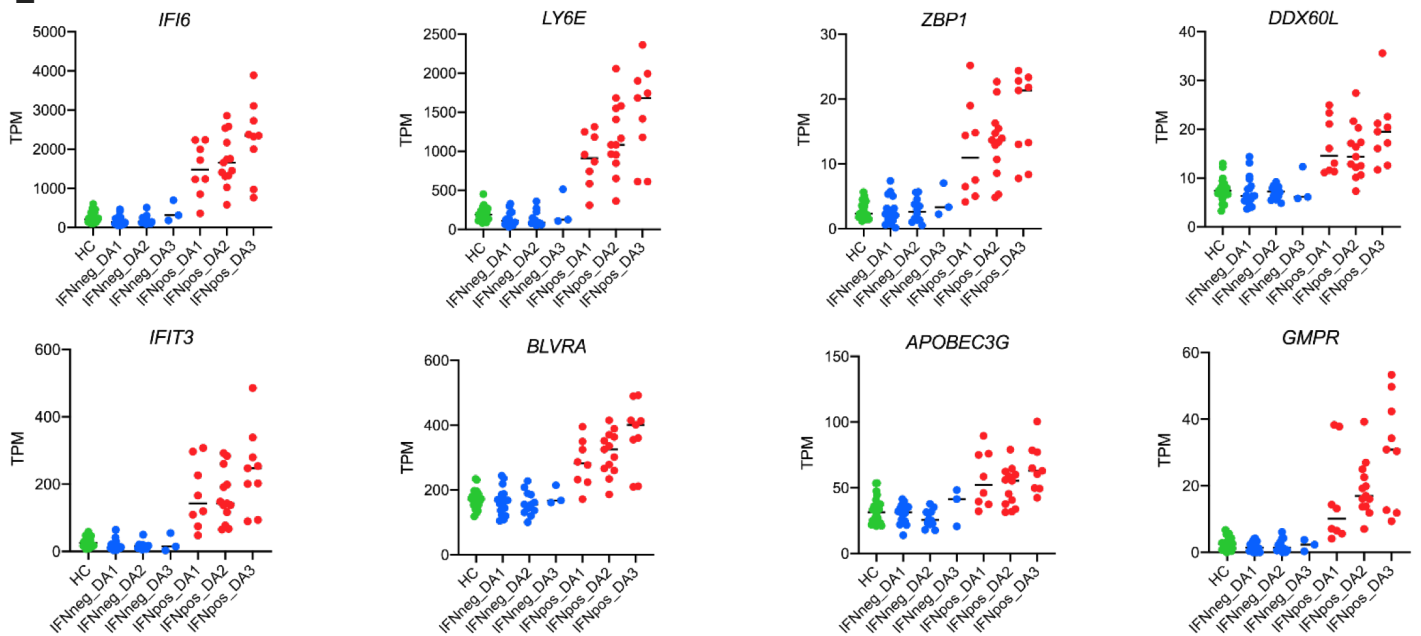
C



D

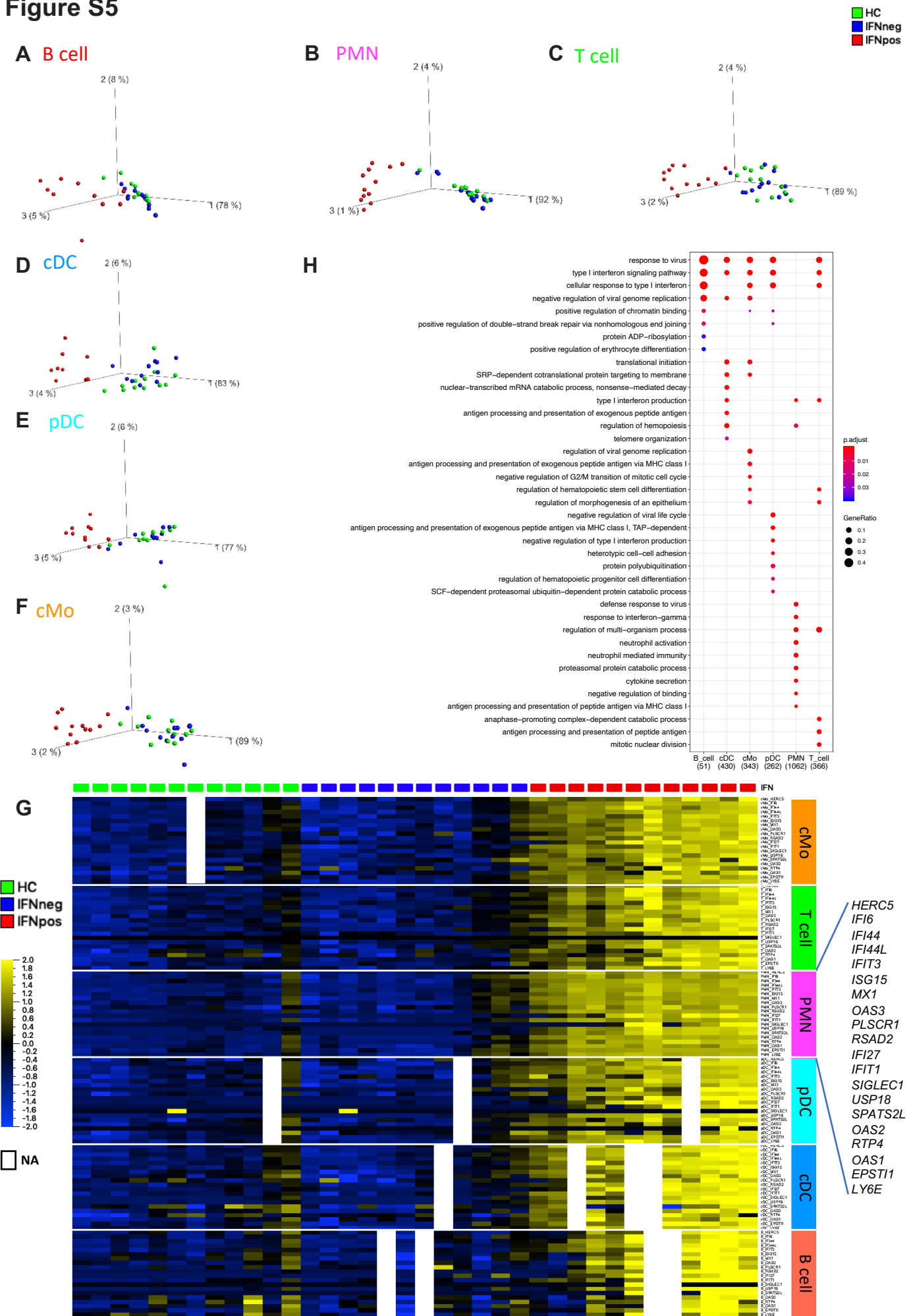


E



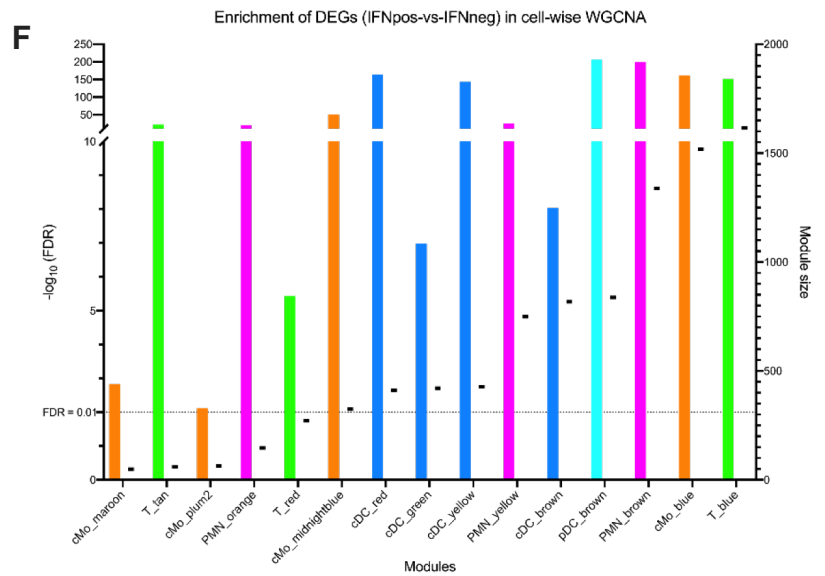
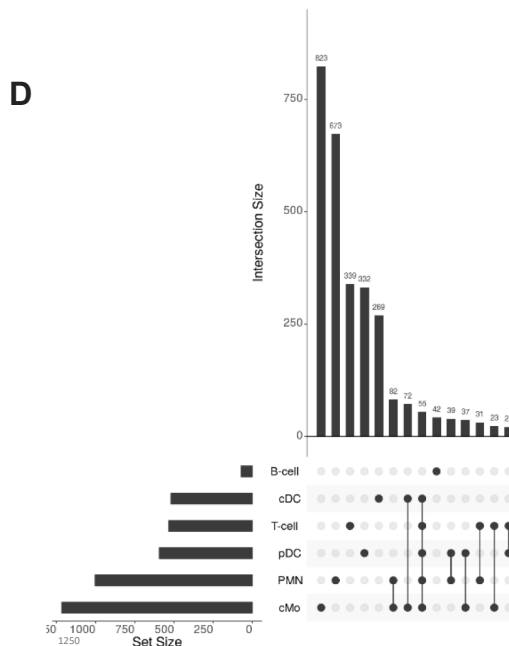
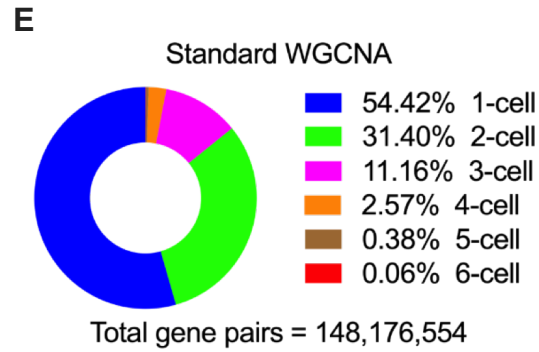
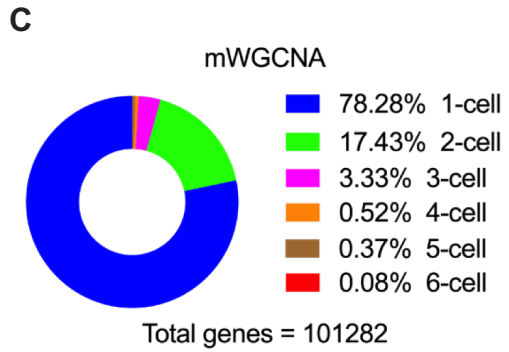
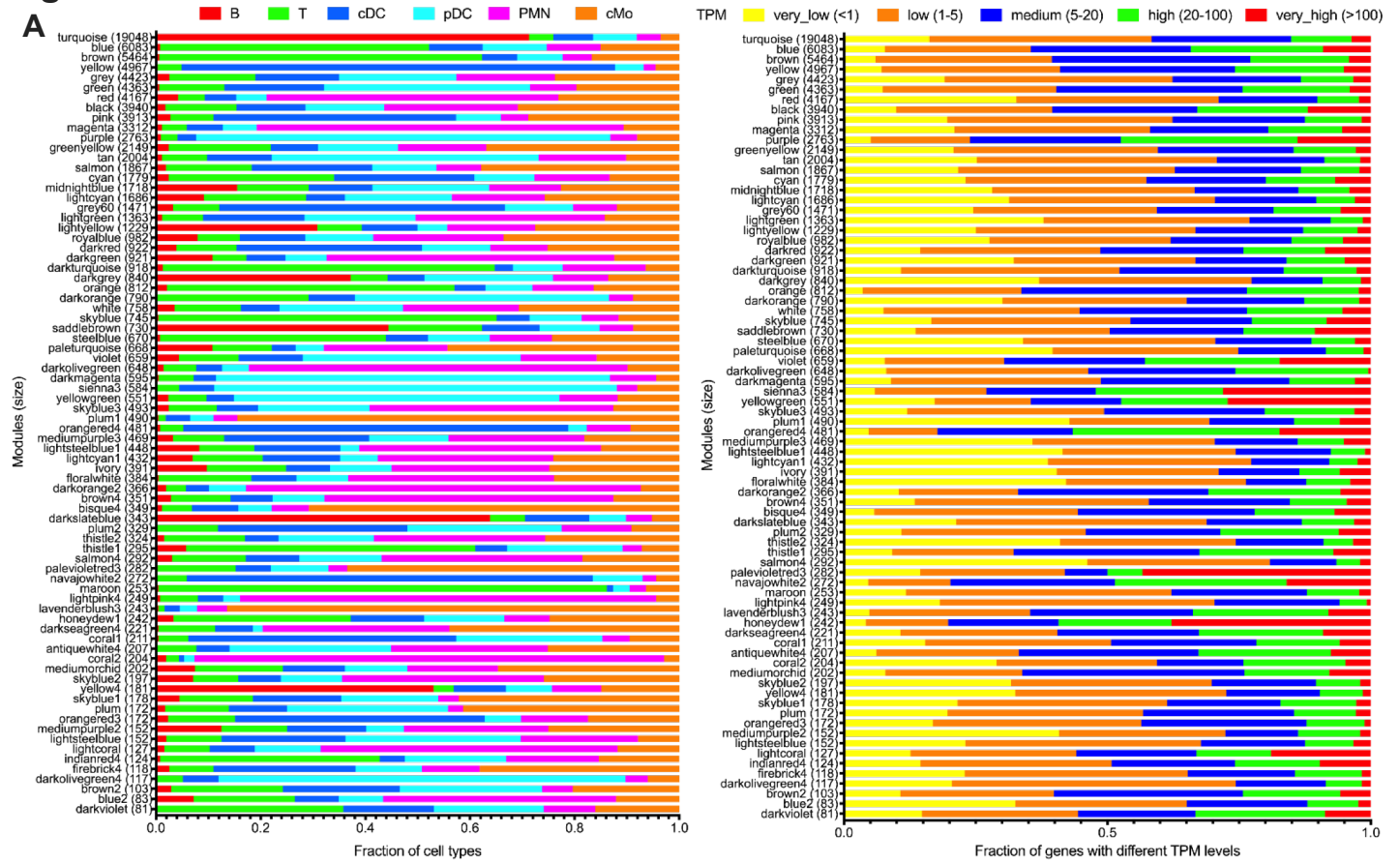
Supplemental Figure S4. Combined analysis of differential network and gene expression of classical monocytes reveals two known immune modulators (*TNFSF13B* and *IL1RN*) and whose expression is dysregulated in SLE. (A) The plot shows both differentially expressed (DEGs) and connected genes (DCGs) from the IFNpos network based green and grey modules only. Where the X-axis is the difference in the connectivity ($\text{DiffK} = K1 - K2$; $K1$ =connectivity in IFNpos network; $K2$ =connectivity in IFNneg network) and the Y-axis is the DESeq2 based fold-change (\log_2). By default, the 'grey' module is generated by WGCNA for non-co-expressed genes so it shows that green module genes are well-expressed as well as well-connected in the IFNpos in comparison to grey module genes. **(B)** Functional annotations (generated by clusterProfiler) from all green module genes in IFNpos network. The color shows the significance (in terms of P_{adj}), the size is gene counts in annotation, and the X-axis shows gene ratio. **(C)** Expression changes of *TNFSF13B* (*BAFF*) and *IL1RN* in longitudinal data from patients S1057 and S1053, where genes are plotted on different Y-axes; *TNFSF13B* on Y1-axis in red and *IL1RN* on Y2-axis in blue. The lower panel of this plot shows corresponding IFN response status over multiple longitudinal visits. **(D)** The DICE database-based expression of *TNFSF13B* and *IL1RN* in different immune cell types from healthy donors. Where X-axis is expression (TPM) level and Y-axis shows different immune cell types. **(E)** The individual expression plot of genes of interest from the green module using IFN response status as well as SLEDAI categories.

Figure S5



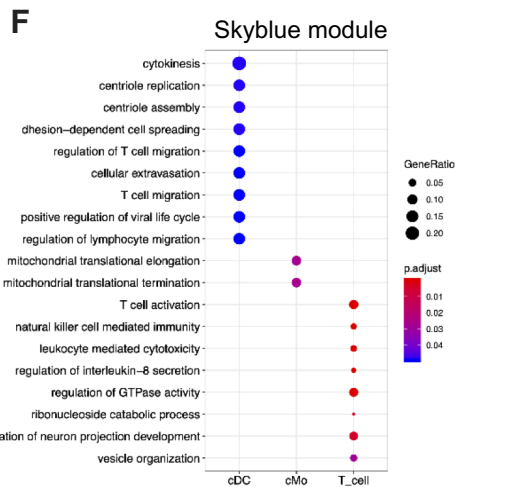
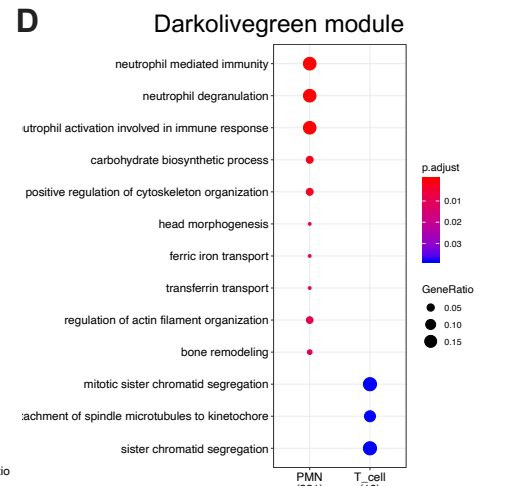
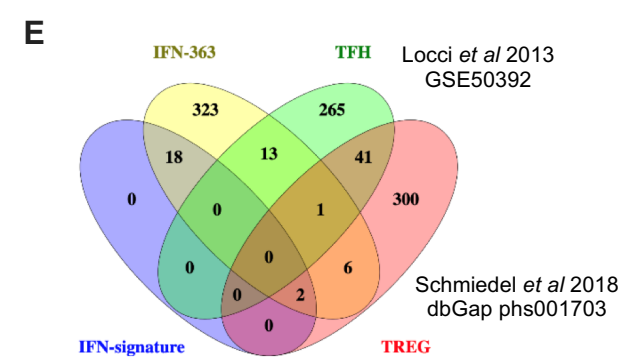
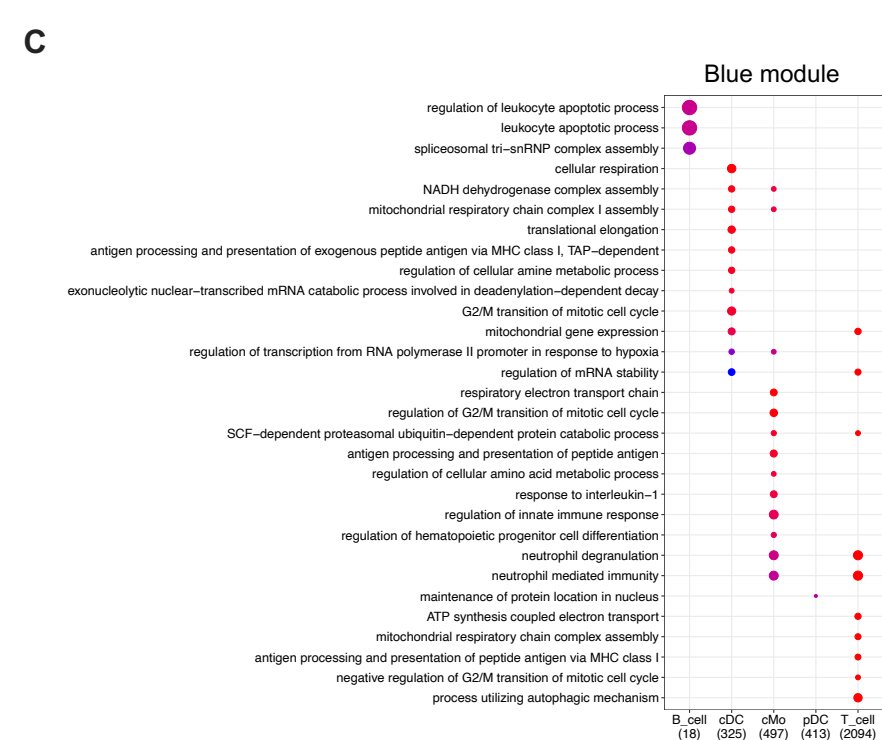
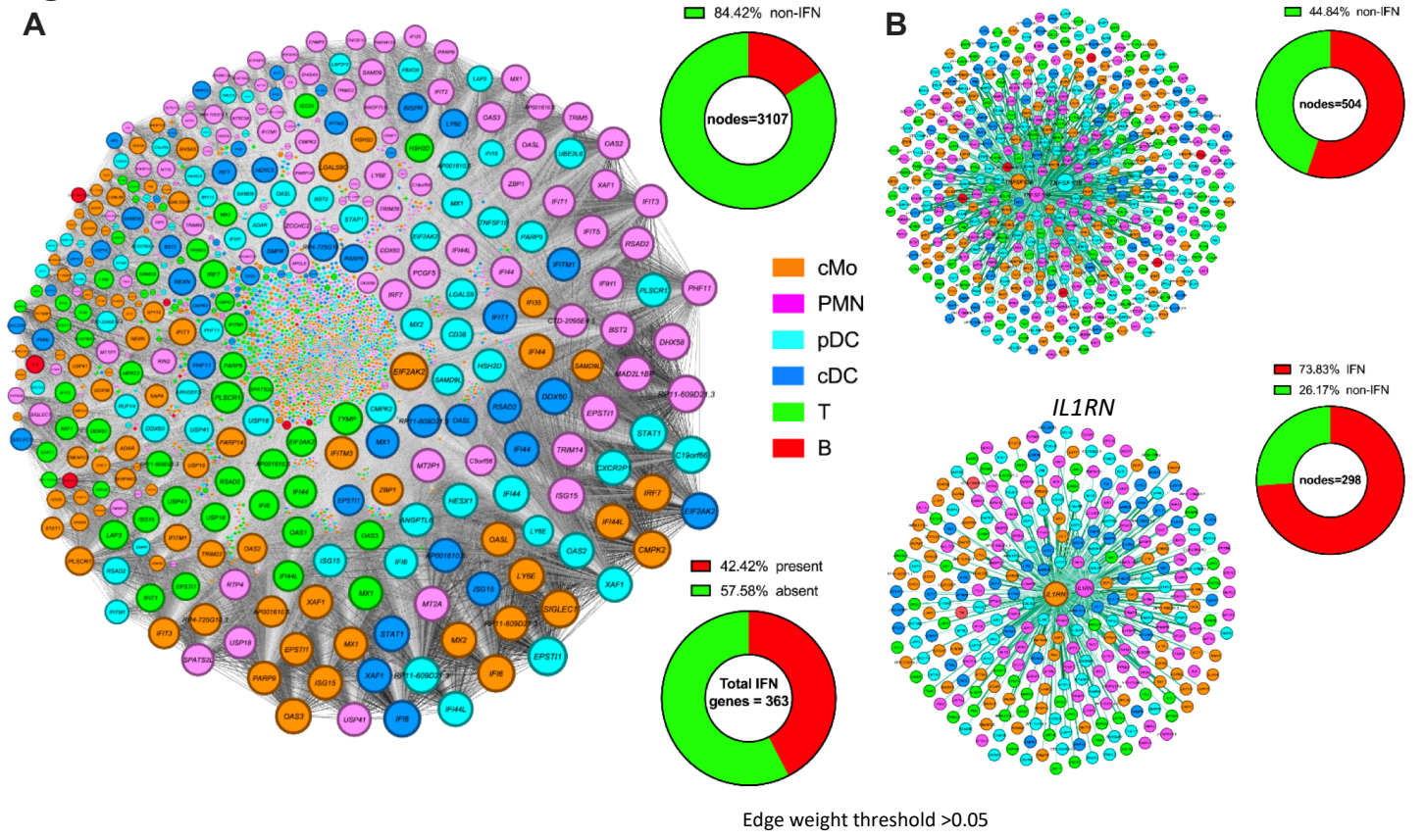
Supplemental Figure S5. IFN-based molecular signature is independent of immune cell types. **(A-F)** PCA plots of different cell types based on IFN signature genes (IFN-20). **(G)** Expression of IFN-20 genes presented as row-wise z-scores of TPM in IFNpos (red), IFNneg (blue) and HC (green) for each cell type separately. The expression profile is absent (used blank or 'NA') for a few patients in some cell types. Patients were clustered within their IFN status based on PC1 values. **(H)** Functional annotations (generated by clusterProfiler) of all DEGs (IFNpos-vs-IFNneg) from different cell types. The color shows the significance (in terms of P.adj) and the size is gene ratio of annotations.

Figure S6

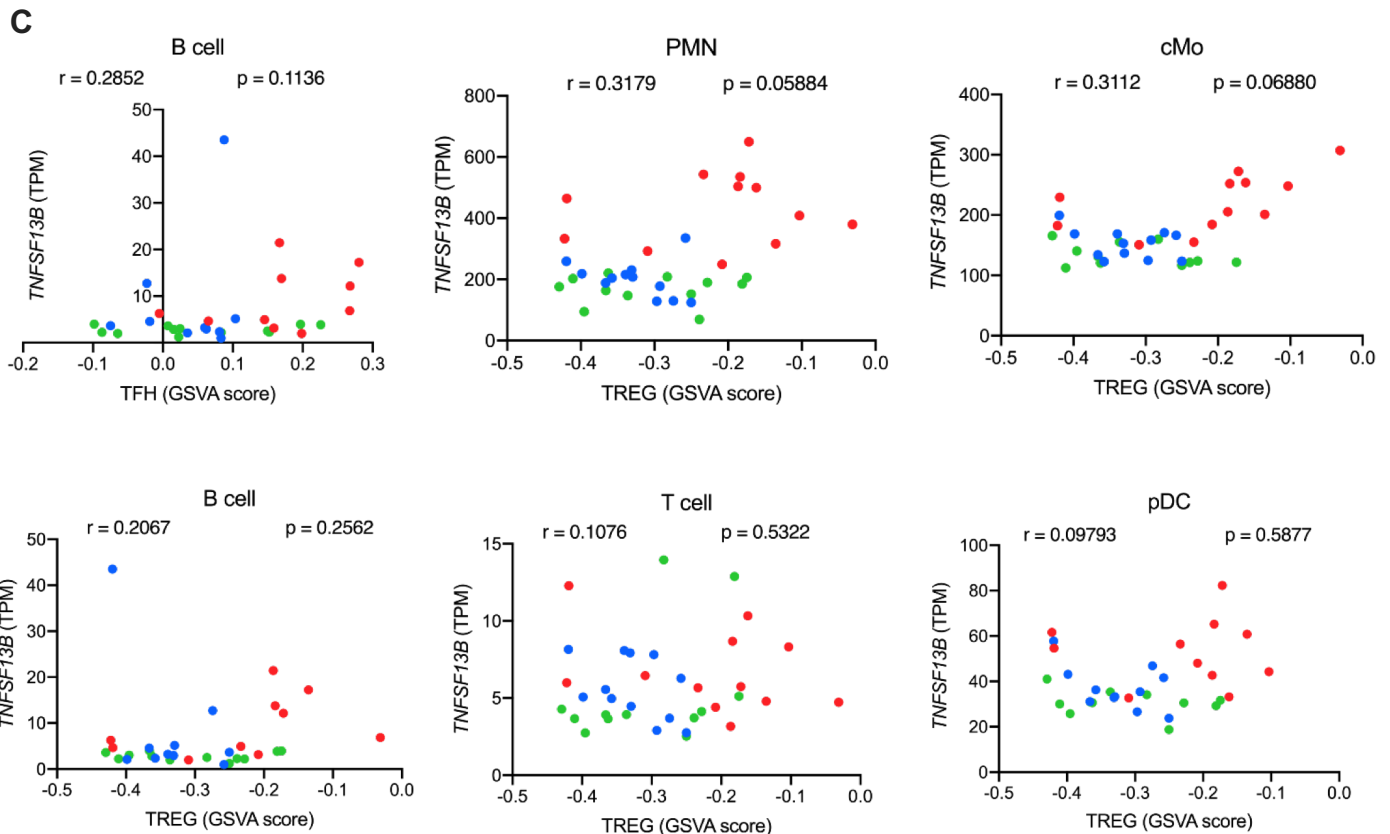
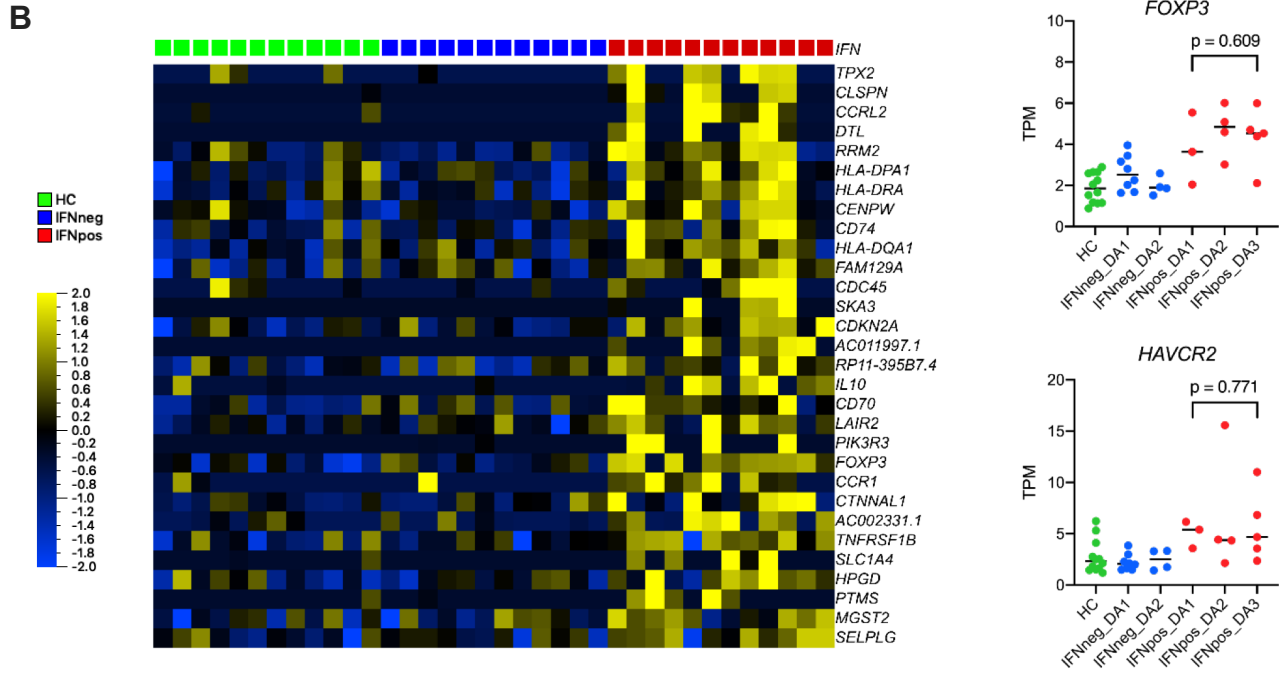
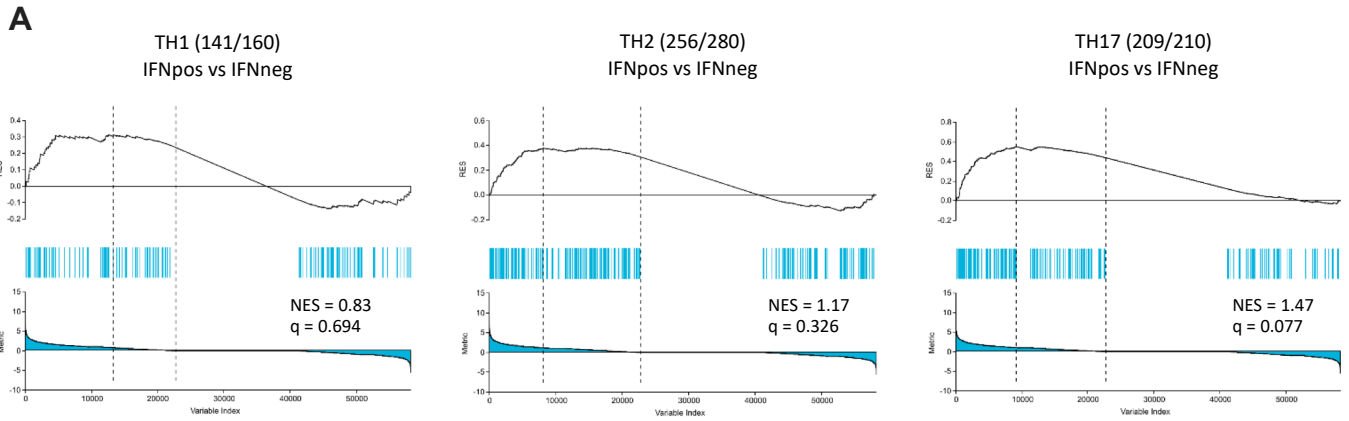


Supplemental Figure S6. Proportion of multi-genes in mWGCNA and gene-pairs in standard WGCNA from different cell types. **(A)** The figure depicts the mWGCNA analysis using combined transcriptomes from six different cell types. The mWGCNA has generated 78 different modules and a stacked column graph shows the proportion of genes from 6 different cell types in each module. The Y-axis shows all modules and the X-axis is the fraction of total number of genes in the corresponding modules. Gene numbers from different cell types are displayed in different colors as depicted in the cartoon. **(B)** The figure depicts the different gene expression levels (based on TPM) in mWGCNA analysis using combined transcriptomes from six different cell types. The mWGCNA has generated 78 different modules and a stacked column graph shows the proportion of genes from 5 different expression levels (very high, high, medium, low, and very low) in each module. The Y-axis shows all modules and the X-axis is the fraction of genes with different TPM levels in the corresponding modules. Different expression levels are displayed in different colors as depicted in the cartoon. **(C)** The donut chart shows the proportion of number of genes from different cell-types in mWGCNA. In this 78.28%, 17.43%, 3.33%, 0.52%, 0.37%, and 0.08% genes are coming from one, two, three, four, five, and six different cell types respectively. **(D)** The UpSetR plot shows the proportion of number of genes from different cell-types in the black module of mWGCNA. We only show sets with a minimum of 20 genes. **(E)** The donut chart shows the proportion of different gene-pairs in standard WGCNA from different cell types. Where 54.42%, 31.40%, 11.16%, 2.57%, 0.38%, and 0.06% gene pairs come from one, two, three, four, five, and six different cell types respectively. **(F)** Plot of selected 15 modules (from standard WGCNA of cell-types) that have significant enrichment of DEGs (IFNpos-vs-IFNneg) for that particular cell type. The X-axis shows modules with their enrichment in each cell type and Y1-axis shows $-\text{Log}_{10}$ value of FDR (based on hypergeometric test). Modules are sorted based on their increasing module size (shown as black dash on the Y2-axis).

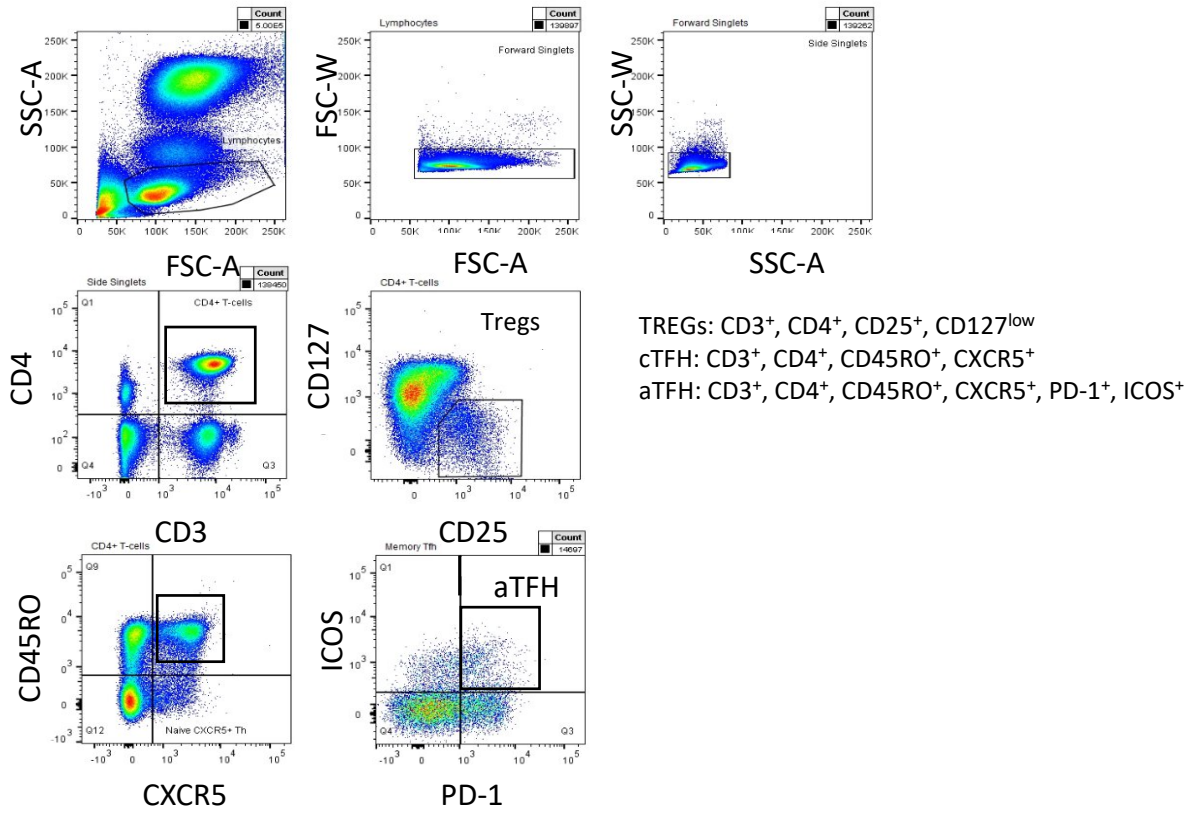
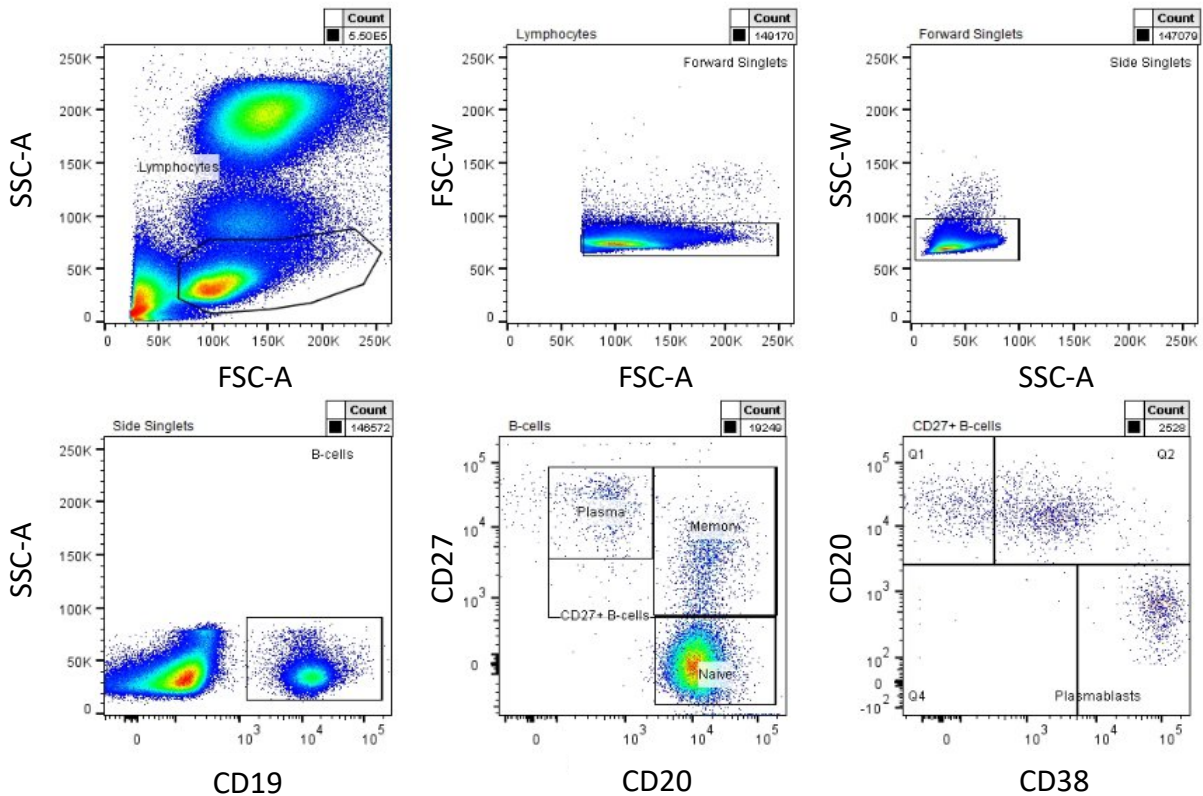
Figure S7



Supplemental Figure S7. Multi cell type WGCNA (mWGCNA) reveals IFN-driven cross-talk between T cells and B cells. (A) The network of the black module genes visualized by Gephi. The nodes are colored according to cell of origin and sized according to the number of edges (connections), and the edge thickness is proportional to the strength of co-expression. The top donut chart shows that 15.58% nodes are IFN-related genes (from IFN-363) and the bottom donut plot shows that 42.42% of IFN-363 genes, from at least one-cell type, are present in the black module. **(B)** The Gephi based visualization of *TNFSF13B* (*BAFF*) (top) and *IL1RN* (bottom) connected genes in the black module and the corresponding donut chart show the proportion of IFN and non-IFN related nodes (based on IFN-363 genes). *TNFSF13B* from three different cell types (cMo, PMN, and pDC) and *IL1RN* from cMo and PMN are present in the black module. **(C-E)** Functional annotations (generated by clusterProfiler) of genes from each cell type in the blue **(C)**, darkolivegreen **(D)** modules. The color shows the significance (in terms of P.adj) and the size is gene ratio of annotations. **(E)** A Venn diagram shows the overlap of TFH and TREG signature genes with IFN-20 and IFN-363 genes. We only used genes that are exclusively present in TFH (265) and TREG (300) signatures. **(F)** Similar plot to **(C-D)** but for the skyblue module.

Figure S8

Supplemental Figure S8. TFH and TREG signatures are more enriched in IFNpos patients and *TNFSF13B* expression in multiple cell types correlates with TFH feature. (A) GSEA of TH1 (left), TH2 (middle), and TH17 (right) in the transcriptome of IFNpos versus IFNneg in T cells, presented as running enrichment score (RES) for the gene set, from most over-represented genes at left to most under-represented at right; values above the plot represent the normalized enrichment score (NES) and false discovery rate (FDR)-corrected significance value; Kolmogorov-Smirnov test. **(B)** A heatmap (left) shows expression of the top 30 TREG-related genes, where each gene is presented as row-wise z-scores of TPM values in IFNpos (red), IFNneg (blue) and HC (green); each column represents an individual patient. Individual expression plots (right) of two TREG related genes (*FOXP3* and *HAVCR2*) using IFN response status as well as SLEDAI categories. The provided p-values are based on unpaired *t*-test. **(C)** The correlation between GSVA score of TFH and TREG gene set and expression (TPM values) of *TNFSF13B* from different cell types. The Spearman's correlation with significance value (p-value) is given.

Figure S9**A Gating Strategy for TREG and TFH****B Gating Strategy for Plasma Cells and Plasmablasts**

Supplemental Figure S9. Flow cytometry gating strategies. (A) Gating scheme of whole blood used for determination of the frequency of TREG, circulating (cTFH), and activated T follicular helper (aTFH) cells. Live lymphocytes were gated based on their forward side scatter versus side scatter profile followed by single cells gating by FSC then SSC and then finally according to their cell surface phenotype as follows. TREGs (CD3⁺, CD4⁺, CD25⁺ CD127^{low}), circulating TFH (cTFH) (CD3⁺, CD4⁺, CXCR5⁺, CD45RO⁺), and activated circulating TFH (aTFH) (CD3⁺, CD4⁺, CXCR5⁺, CD45RO⁺, PD-1⁺, ICOS⁺). **(B)** Gating scheme of whole blood used for determination of the frequency of plasma cells and plasmablasts. Live lymphocytes were gated based on their forward side scatter versus side scatter profile followed by single cells gating by FSC then SSC and then finally according to their cell surface phenotype as follows. Plasma Cells (CD19⁺, CD20⁻, CD27⁺ and plasmablasts (CD19⁺, CD20⁻, CD27⁺, CD38⁺).