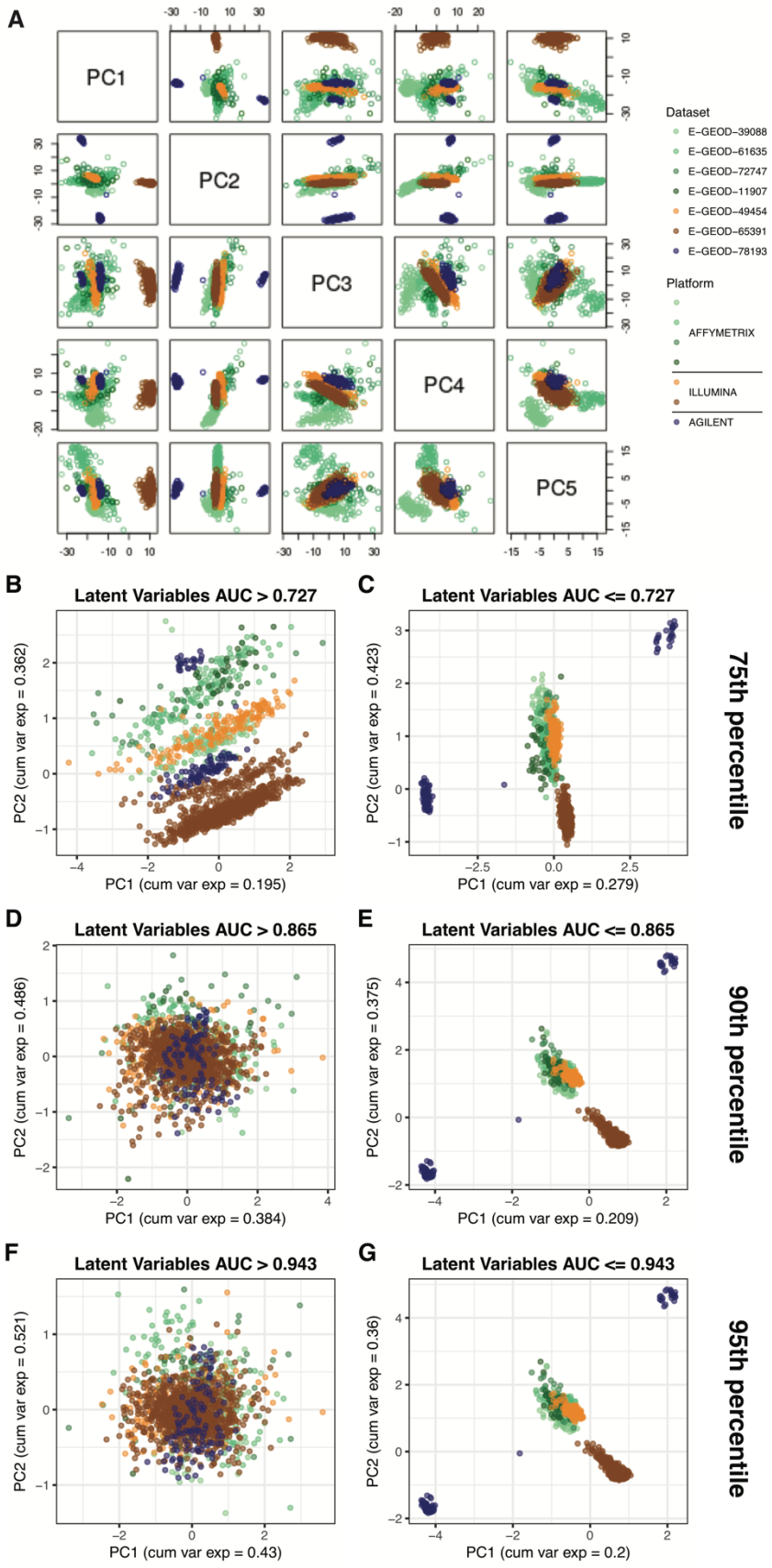

A
PC1
PC2
PC3
PC4
PC5
Dataset
E−GEOD−39088
E−GEOD−61635
E−GEOD−72747
E−GEOD−11907
E−GEOD−49454
E−GEOD−65391
E−GEOD−78193
Platform
AFFYMETRIX
ILLUMINA
AGILENT
B
Latent Variables AUC > 0.727
PC2 (cum var exp = 0.362)
PC1 (cum var exp = 0.195)
C
Latent Variables AUC <= 0.727
PC2 (cum var exp = 0.423)
PC1 (cum var exp = 0.279)
75th percentile
D
Latent Variables AUC > 0.865
PC2 (cum var exp = 0.486)
PC1 (cum var exp = 0.384)
E
Latent Variables AUC <= 0.865
PC2 (cum var exp = 0.375)
PC1 (cum var exp = 0.209)
90th percentile
F
Latent Variables AUC > 0.943
PC2 (cum var exp = 0.521)
PC1 (cum var exp = 0.43)
G
Latent Variables AUC <= 0.943
PC2 (cum var exp = 0.36)
PC1 (cum var exp = 0.2)
95th percentile

Figure S1. Principal components 1-5 for the SLE WB expression compendium. Related to Figure 2. Points are samples. Samples are colored by dataset of origin and datasets from the same platform manufacturer are similar colors. PC1-5 account for 0.502 of the variance explained. Panels B-G display the first two PCs from different subsets of the latent space or B matrix, either those associated with input gene sets (e.g., above the labeled AUC threshold) or those that are not. AUC thresholds were chosen based on the distribution of all AUC values.

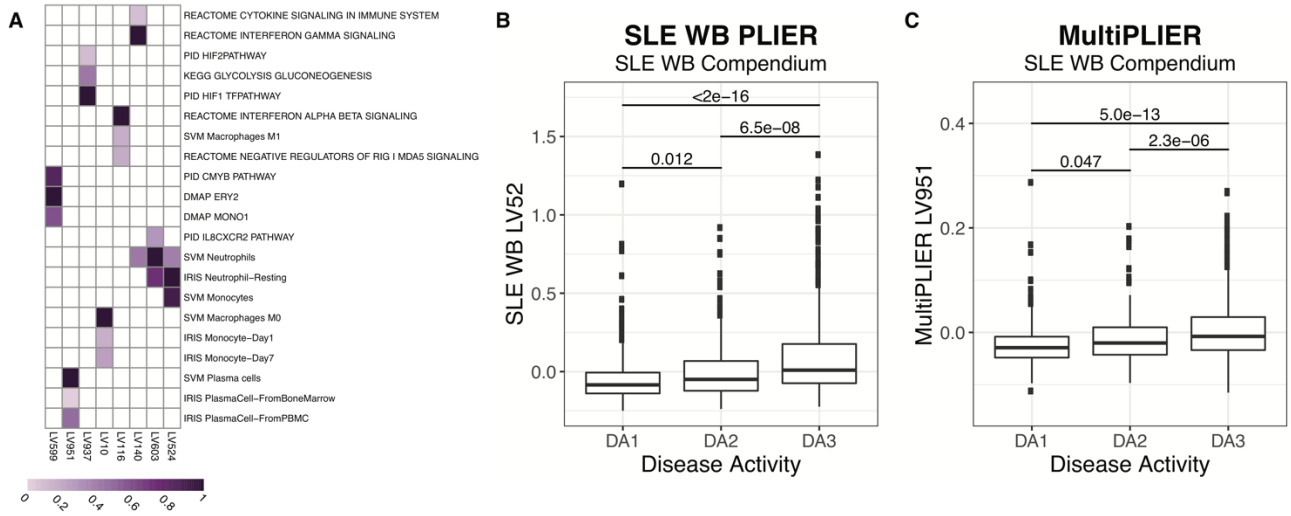


Figure S2. MultiPLIER learns a plasma cell signature that is elevated in more severe disease. Related to Figure 3. (A) Selected latent variables (LVs) from the MultiPLIER U matrix. Purple fill in a cell indicates a non-zero value and a darker purple indicates a higher value. Only pathways with FDR < 0.05 in displayed latent variables are shown. We selected latent variables discussed in the main text. **(B-C)** In Banchereau, et al. citation the authors stratified patients into three disease activity groups—DA1 (SLEDAI: 0-2), DA2 (SLEDAI: 3-7), and DA3 (SLEDAI > 7)—and demonstrated that patients with more severe disease had higher plasmablast counts (via FACS). We did not provide a plasmablast-specific gene set to the PLIER model during training. However, plasma cell from signatures adapted from the Immune Response In Silico (IRIS) project (Abbas et al., 2005) and Newman, et al. (Newman et al., 2015)(included in the PLIER R package). A plasmablast is the precursor to a plasma cell and we therefore assume that these cell types share gene expression patterns. Both the SLE WB **(B)** and MultiPLIER **(C)** models learn a latent variable associated with this signature (SLE WB LV52 and MultiPLIER LV951, respectively). P-values are from a pairwise t-test with Bonferroni correction. Both latent variables show a similar pattern to the original publication FACS plasmablast counts. However, there is evidence that the MultiPLIER latent variable may more specific to plasma cells, as SLE WB LV52 is also significantly associated (FDR < 0.05) with REACTOME Unfolded Protein Response and REACTOME Asparagine N-linked Glycosylation (Fabregat et al., 2017)(Figure 2A). MultiPLIER LV951 is only significantly associated with plasma cell gene sets (data not shown).

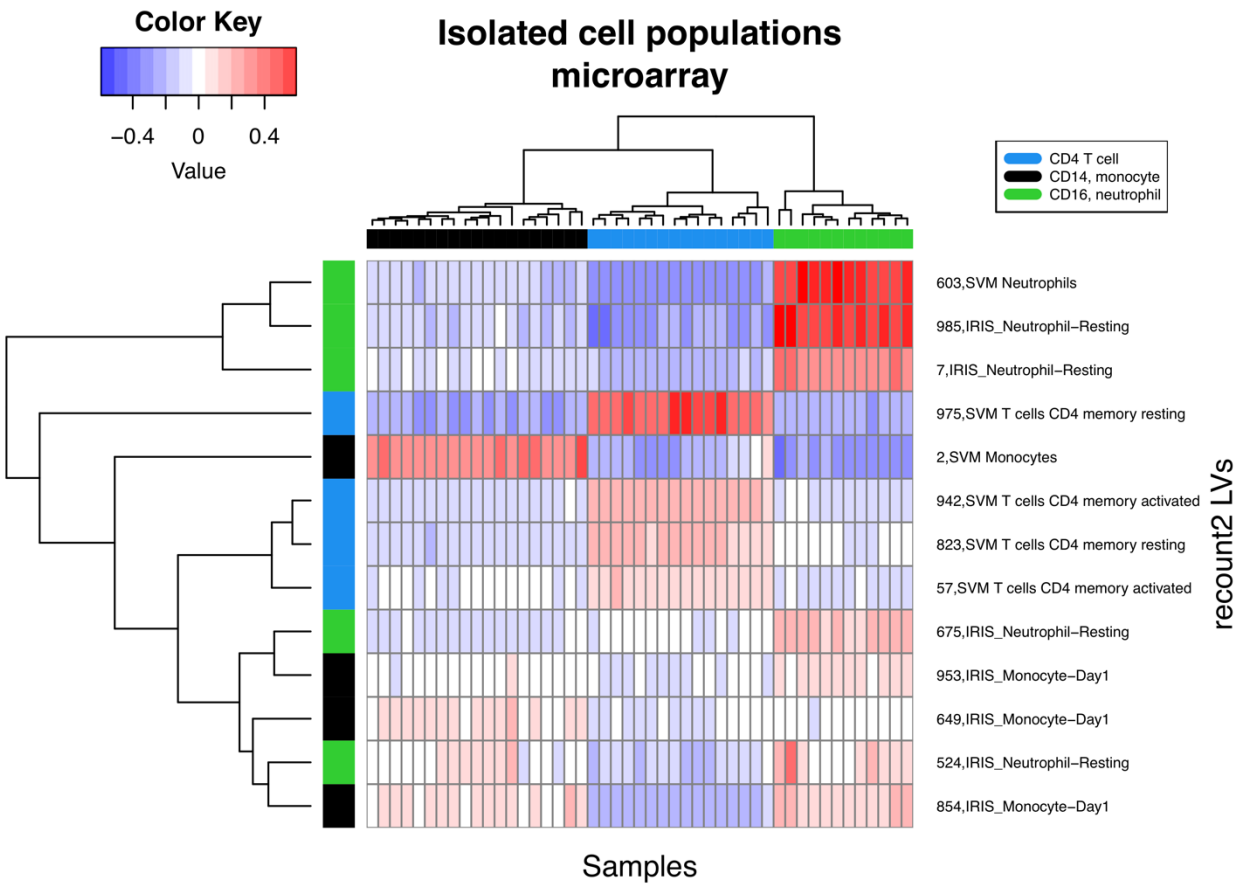
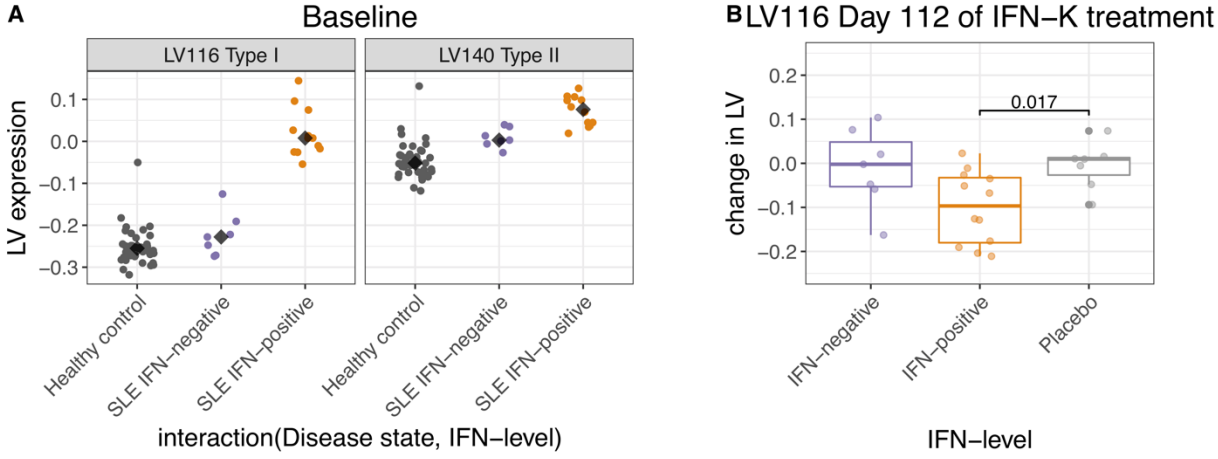
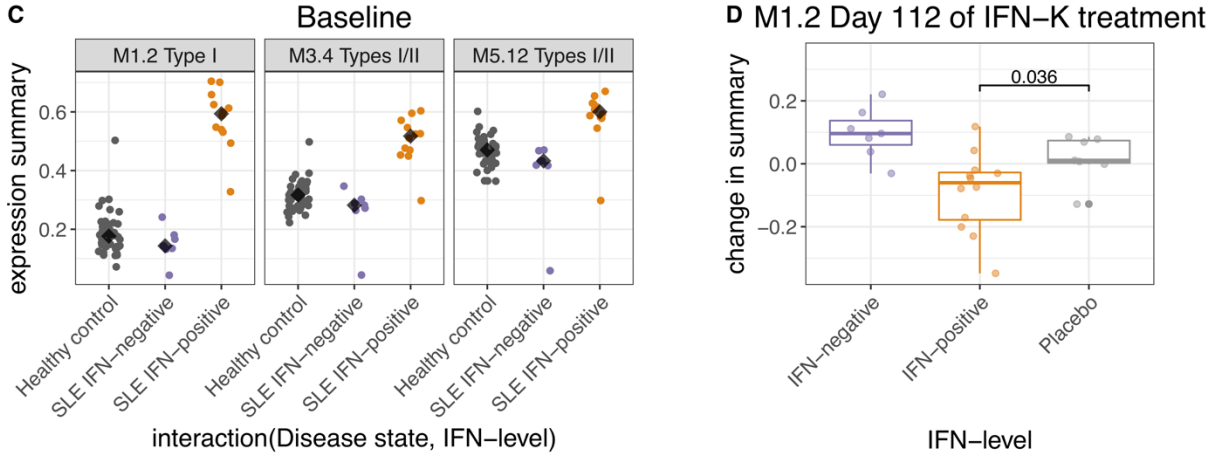


Figure S3. Projection of isolated leukocyte gene expression data, as measured by microarray, into MultiPLIER latent space demonstrates that the model captures additional cell type-specific signatures. Related to Figure 3. Heatmap of a subset of the B matrix for E-MTAB-2452 (McKinney et al., 2015), a dataset comprised of isolated immune subsets ($CD4^+$ T cells, $CD14^+$ monocytes, and $CD16^+$ neutrophils) from patients with autoimmune diseases. Latent variables (rows) were selected if their names contained one of the following terms: CD4, monocyte, or neutrophil. The sample (column) annotation bar is colored based on the cell subset of that sample. The latent variable (row) annotation bar is colored by the term associated with the latent variable.

MultiPLIER



Whole Blood Modular Framework



SLE WB PLIER

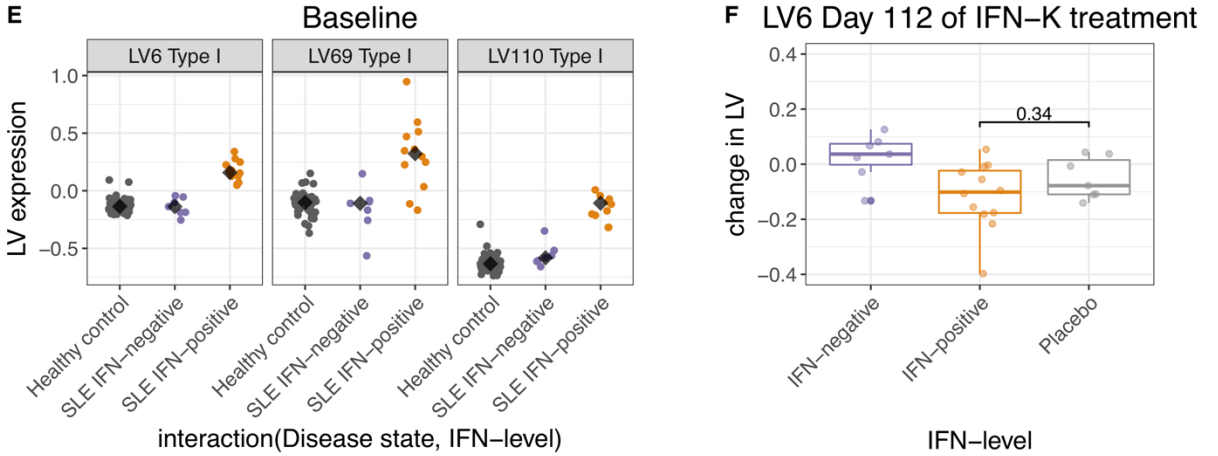
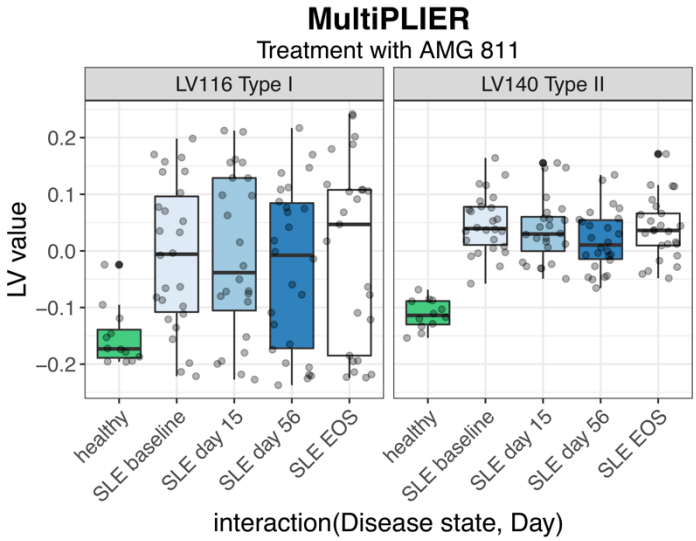
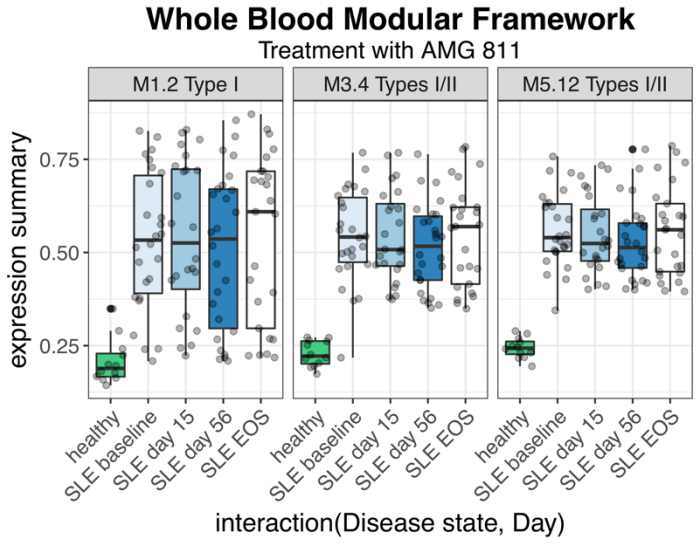


Figure S4. Expression patterns from IFN-K (type I IFN blockade) trial in the MultiPLIER latent space, whole blood modular framework, and SLE WB PLIER latent space. Related to Figure 5. Lauwerys, et al. (Lauwerys et al., 2013) is a clinical trial of the IFN-alpha-kinoid (IFN-K) therapeutic that results in the blockade of type I IFN via polyclonal antibody response. The authors of the original IFN-K publication stratified patients with SLE into two groups: IFN-negative (n = 9), who had IFN-alpha induced gene expression signatures that resembled healthy controls, and IFN-positive patients (n = 18), who had higher expression of the IFN-alpha signature they derived through *ex vivo* stimulation of healthy control blood with IFN-alpha. (Group labels from the original publication were not associated with the accession.) **(A)** Patients can be stratified into IFN-positive and IFN-negative groups using baseline expression of MultiPLIER LV116 (latent variable 116 from the MultiPLIER model). The expression of type II latent variable, MultiPLIER LV140, does not show a clear IFN-negative and IFN-positive group. **(B)** Treated IFN-positive patients show a reduction in MultiPLIER LV116 expression at day 112 as compared to the placebo group (p-value from Wilcoxon rank sum test). Chiche, et al. (Chiche et al., 2014) identified three IFN-related modules—M1.2, M3.4, and M5.12—and demonstrated that M1.2 was strongly induced by type I IFN, whereas the other two modules were similarly responsive to type I and type II IFN. **(C)** Type I module M1.2, and to a lesser extent M3.4, from the whole blood modular framework divides patients into IFN-positive and IFN-negative groups. Expression for each module was summarized using the mean expression value of all genes in the module. **(D)** Treated IFN-positive patients show a reduction in M1.2 expression at day 112 as compared to the placebo group (p-value from Wilcoxon rank sum test). **(E)** LV6, LV69, and LV110 from the SLE WB PLIER model were all significantly associated with type I IFN (REACTOME Interferon alpha/beta signaling). Patients were classified as IFN-negative if they had one of the nine lowest expression values in 2 out of 3 of these latent variables (typically LV6 and LV110). Recall that these samples are in the training data for this model. **(F)** Treated IFN-positive patients do not show a decrease in LV6 expression at day 112 as compared to placebo (p-value from Wilcoxon rank sum test). SLE WB LV6 was selected for display and comparison as it had the highest AUC for REACTOME Interferon alpha/beta signaling in the SLE WB model. Displayed p-values are not corrected for multiple hypothesis testing.


MultiPLIER
Treatment with AMG 811
LV116 Type I
LV140 Type II
LV value
0.2
0.1
0.0
−0.1
−0.2
healthy
SLE baseline
SLE day 15
SLE day 56
SLE EOS
interaction(Disease state, Day)


Whole Blood Modular Framework
Treatment with AMG 811
M1.2 Type I
M3.4 Types I/II
M5.12 Types I/II
expression summary
0.75
0.50
0.25
healthy
SLE baseline
SLE day 15
SLE day 56
SLE EOS
interaction(Disease state, Day)

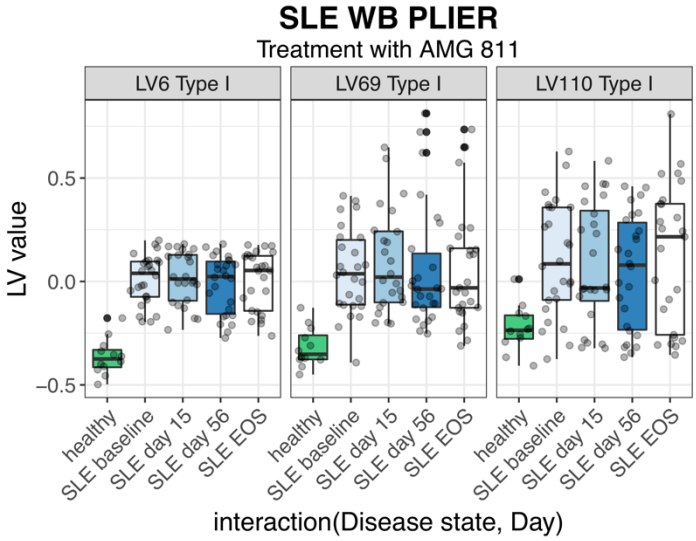

SLE WB PLIER
Treatment with AMG 811
LV6 Type I
LV69 Type I
LV110 Type I
LV value
0.5
0.0
−0.5
healthy
SLE baseline
SLE day 15
SLE day 56
SLE EOS
interaction(Disease state, Day)

Figure S5. Expression patterns from trial of type II IFN blockade agent AMG 811 in the MultiPLIER latent space, whole blood modular framework, and SLE WB PLIER latent space. Related to Figure 5. The type II IFN latent variable (LV) from the MultiPLIER model, LV140, shows a trend towards decreased expression during days 15 and 56 of the AMG811 trial (top panel). The modules most likely to capture type II IFN expression, M3.4 and M5.12, show a similar trend (middle pattern). The latent variables or modules indicative of type I signaling show no such pattern (MultiPLIER LV116, whole blood modular framework M1.2, and all SLE WB latent variables). No SLE WB PLIER latent variable was significantly associated with type II (as determined by the lack of association with the REACTOME Interferon gamma signaling pathway).

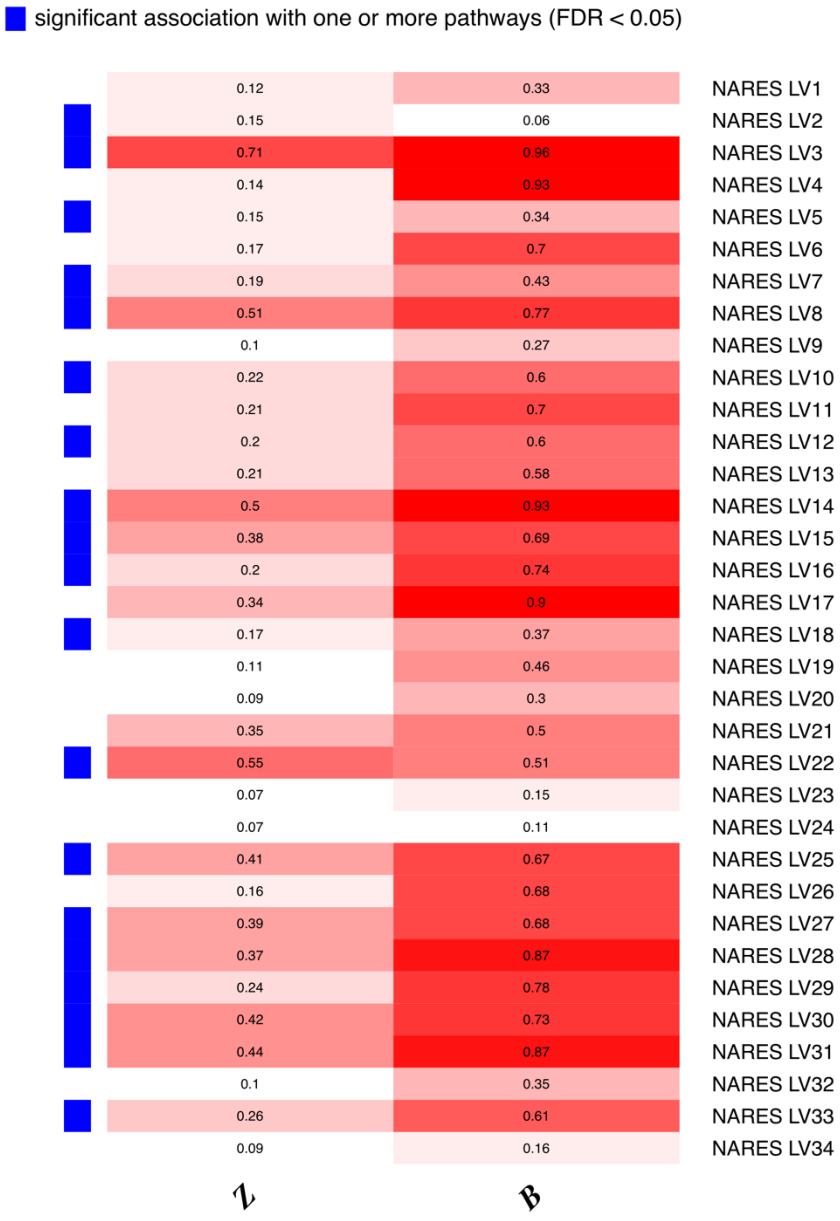


Figure S6. Comparison of the loadings (Z) and latent space (B) of the NARES PLIER model to the MultiPLIER loadings and the projection of the NARES dataset into the MultiPLIER latent space. Related to Figure 6. Heatmap of correlation values between the loadings and latent space of best match latent variables, where rows are the latent variables (LVs) from the NARES PLIER model. A more saturated red color indicates a higher correlation value. A NARES latent variable's best match is determined by calculating the Pearson correlation between loadings (Z matrices) and selecting the latent variable from the MultiPLIER model with the highest correlation to that NARES latent variable. The columns are correlation values for the loadings (Z) and the latent variable expression (B) between that NARES latent variable (row) and its best match in the MultiPLIER model. The row annotation bar indicates whether or not a NARES latent variable is significantly associated (FDR < 0.05) with one or more gene sets supplied during training.

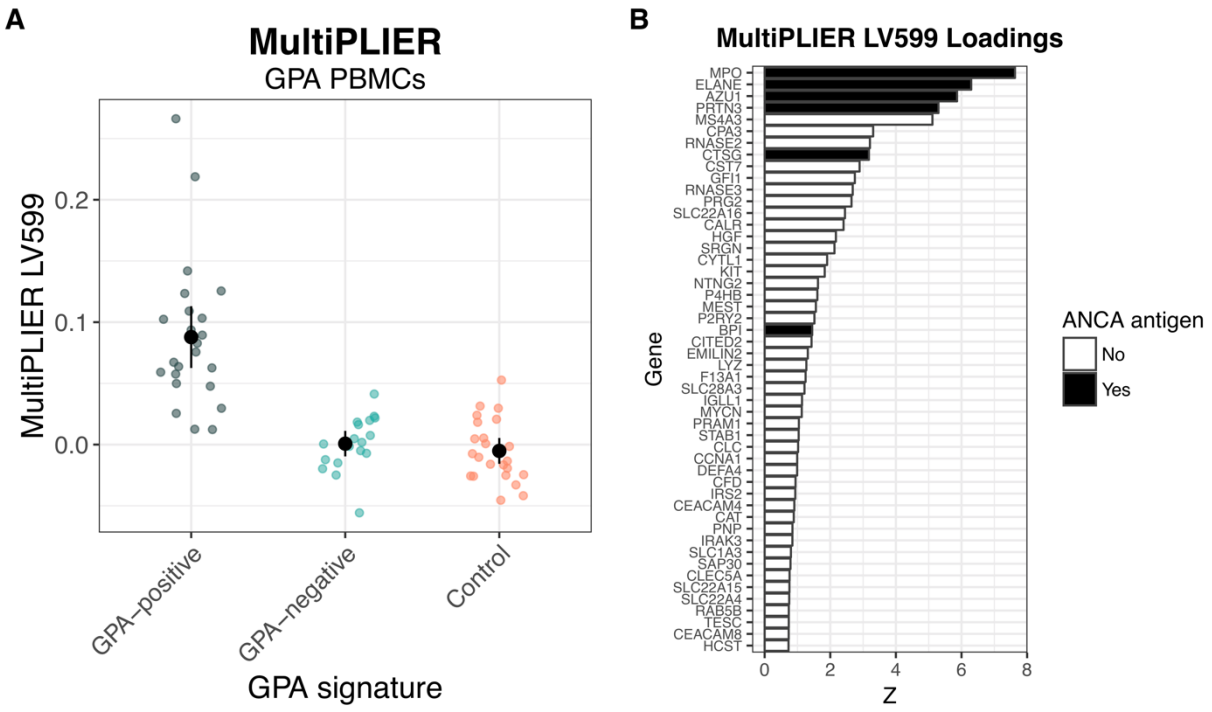


Figure S7. MultiPLIER learns a latent variable (LV) that captures the activity of antigens in antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis despite no vasculitis training data. Related to Figure 7. The peripheral blood mononuclear cell (PBMC) fraction data from Cheadle, et al. (Cheadle et al., 2010) was projected into the MultiPLIER latent space. The authors of Cheadle, et al. identified three groups: GPA-positive ("WG-positive" for Wegener's Granulomatosis in the original 2010 publication), GPA-negative ("WG-negative" in the original publication), and controls. **(A)** MultiPLIER LV599 is differentially expressed between the three groups (FDR = 1.17e-08). Points and bars represent mean ± 2 * SEM. **(B)** The loadings of the top (highest weight) 50 genes for MultiPLIER LV599 includes ANCA antigens, shown with black bars. The original publication reported differential expression of ANCA antigens in the PBMC fraction.

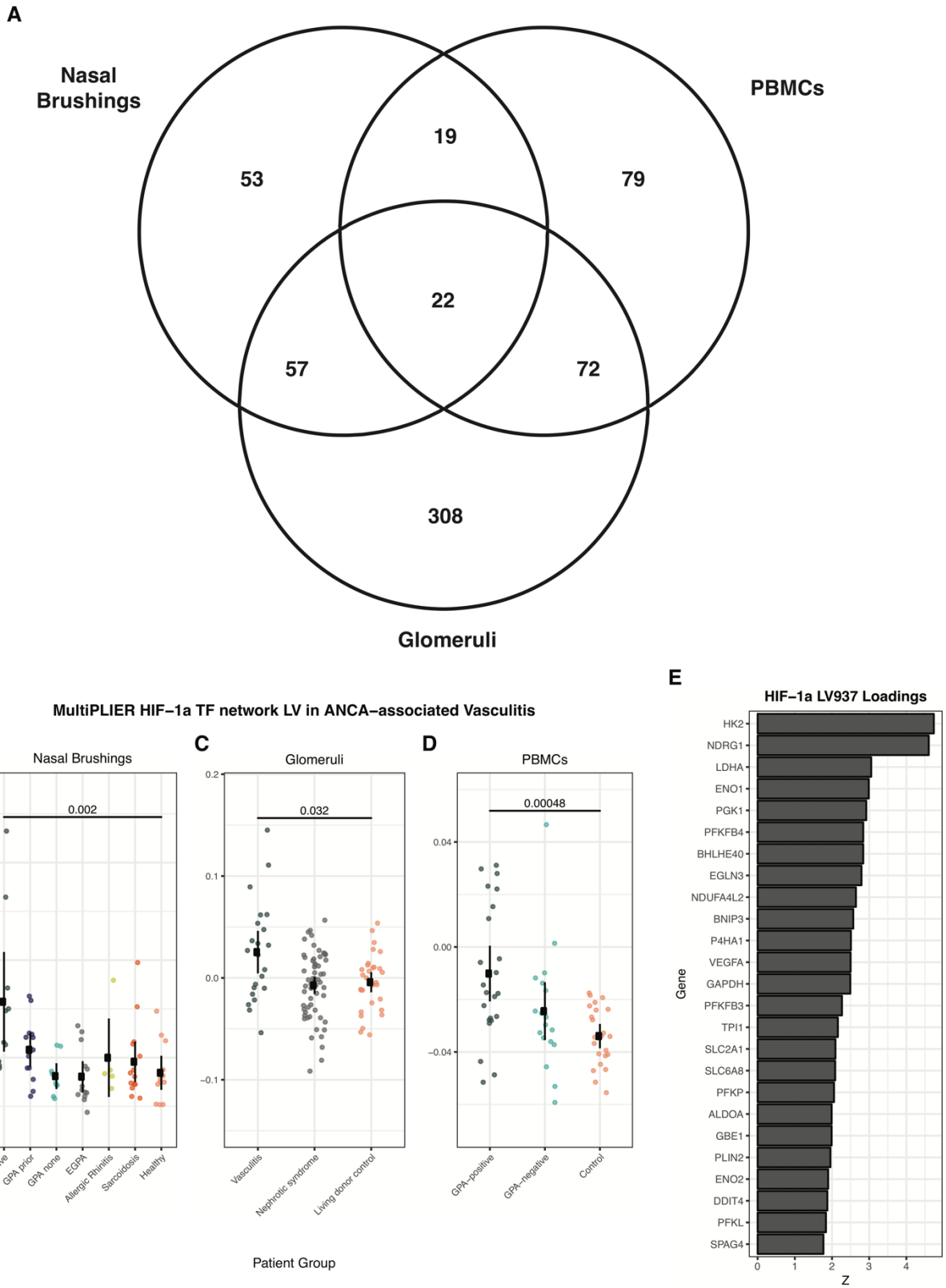

A
Nasal Brushings
PBMCs
Glomeruli
53
19
79
22
57
72
308
MultiPLIER HIF−1a TF network LV in ANCA−associated Vasculitis
B
Nasal Brushings
0.002
MultiPLIER LV937
GPA active
GPA prior
GPA none
EGPA
Allergic Rhinitis
Sarcoidosis
Healthy
C
Glomeruli
0.032
Vasculitis
Nephrotic syndrome
Living donor control
D
PBMCs
0.00048
GPA−positive
GPA−negative
Control
Patient Group
E
HIF−1a LV937 Loadings
HK2
NDRG1
LDHA
ENO1
PGK1
PFKFB4
BHLHE40
EGLN3
NDUFA4L2
BNIP3
P4HA1
VEGFA
GAPDH
PFKFB3
TPI1
SLC2A1
SLC6A8
PFKP
ALDOA
GBE1
PLIN2
ENO2
DDIT4
PFKL
SPAG4
Gene
Z

Figure S8. A MultiPLIER-learned latent variable associated with the HIF-1a transcription factor network is differentially expressed in three tissues from ANCA-associated vasculitis (AAV) and shows increased expression in severe or active disease. Related to Figure 7. Differentially expressed latent variables (LVs) were identified by comparing all patient groups and using Benjamini-Hochberg correction (FDR). Latent variables with FDR < 0.05 were considered to be differentially expressed. MultiPLIER LV937 had an FDR < 0.05 in all three cohorts. **(A)** Venn Diagram of latent variables differentially expressed in three tissues. **(B-D)** Jitter plots of MultiPLIER LV937 in three different tissues: nasal brushings (NARES dataset), kidney microdissected glomeruli, and peripheral blood mononuclear cells. Points and bars represent mean ± 2 * SEM. P-values are from a Wilcoxon rank sum test comparing the control group to the AAV group considered to have the most severe or active disease in the cohort. **(E)** The loadings of the top (highest weight) 25 genes for MultiPLIER LV937.

TABLES

Table S1. Terms used to subset recount2 into training sets for the biological context experiments and the set sample sizes. Related to STAR Methods. Labels are from MetaSRA (Bernstein et al., 2017).

Training set	Term or strategy used	Sample size (n)
Blood	UBERON:0000178	3862
Cancer	EFO:0000311	8807
Tissue	MetaSRA sample type = tissue	12396
Cell line	MetaSRA sample type = cell line	14532
Other tissues	Everything but samples that mapped to blood training set	32984

Table S2. Gene sets used for pathway separation evaluations. Related to Figure 5A and STAR Methods.

Evaluation	Set 1	Set 1 gene sets	Set 2	Set 2 gene sets
Interferon (IFN)	Type I IFN	REACTOME Interferon Alpha/Beta Signaling	Type II IFN	REACTOME Interferon Gamma Signaling
Myeloid	Neutrophil	DMAP GRAN3, IRIS Neutrophil-Resting, SVM Neutrophils	Monocyte/macrophage	IRIS Monocyte-Day0, IRIS Monocyte-Day1, IRIS Monocyte-Day7, SVM Monocytes, SVM Macrophages M0, SVM Macrophages M1, SVM Macrophages M2
Proliferation	G1 phase	REACTOME G1 S Transition, REACTOME M G1 Transition, REACTOME, APC C CDH1 Mediated Degradation of CDC20 and other APC C CDH1 Targeted Proteins in Late Mitosis Early G1, REACTOME Cyclin E Associated Events During G1 S Transition, REACTOME G1 Phase, REACTOME Mitotic M M G1 Phases, REACTOME P53 Dependent G1 DNA Damage Response, REACTOME G1 G1 S Phases, Reactome P53 Independent G1 S DNA Damage Checkpoint	G2 phase	REACTOME Mitotic G2 G2 M Phases, REACTOME G2 M Checkpoints